

Karsten Drescher

Asymptotische Eigenschaften von Schätzern in
verallgemeinerten Produkt-Binomial-Verteilungsmodellen
für Kontingenztafeln

Dissertation

zur Erlangung des Grades eines Doktors
der Naturwissenschaften
der Universität Bremen

(Januar 1984)

1. Gutachter: Prof. Dr. Gerhard Osius
2. Gutachter: Prof. Dr. Hans-Peter Kinder





Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41543123_0016

Karsten Drescher

**Asymptotische Eigenschaften von Schätzern in
verallgemeinerten Produkt-Binomial-Verteilungsmodellen
für Kontingenztafeln**

Dissertation

zur Erlangung des Grades eines Doktors
der Naturwissenschaften
der Universität Bremen

(Januar 1984)

1. Gutachter: Prof. Dr. Gerhard Osius
2. Gutachter: Prof. Dr. Hans-Peter Kinder

Für die vielfache Unterstützung beim Erstellen
dieser Arbeit danke ich Prof. Dr. G. Osius.

Inhaltsverzeichnis

1. Einleitung	2
I. KONTINGENZTAFELN MIT ENDLICH VIELEN ZELLEN	
2. Modell	6
3. Beispiele	7
4. Schätzmethoden	10
5. Konsistenz	14
6. Asymptotische Normalität	26
7. Testen von Hypothesen	40
II. KONTINGENZTAFELN MIT UNENDLICH VIELEN ZELLEN	
8. Modell	50
9. Beispiele	53
10. Schätzmethoden	55
11. Konsistenz	56
12. Asymptotische Normalität	78
13. Konsistenz und asymptotische Normalität bei zufällig verteilter Stichprobenanzahl	93
ANHANG A: Mannigfaltigkeiten im $\mathbb{R}^k$	98
ANHANG B: Konvergenzarten bei nicht notwendig meßbaren Variablen	102
Bezeichnungen	107
Literaturverzeichnis	109



1. Einleitung

Gegenstand dieser Arbeit ist die Untersuchung von asymptotischen Eigenschaften von Schätzern bei verallgemeinerten Produkt - Binomial - Verteilungsmodellen. Diese Modelle spielen bei der Analyse von gruppierten Lebenszeiten unter Einschluß von zensierten Daten eine wichtige Rolle. Wir betrachten hierzu das folgende diskrete Lebenszeitmodell:

Bei jeder Versuchseinheit (Individuum) in einer Studie werde eine (ein- oder mehrdimensionale) Kovariable X beobachtet. Jeder Versuchseinheit seien außerdem zwei zufällige Zeiten T_1^* und T_2^* zugeordnet, wobei T_1^* die sogenannte Failurezeit (Lebenszeit, wenn nur das interessierende Risiko vorhanden ist) und T_2^* die sogenannte Censorzeit (Lebenszeit, wenn nur die konkurrierenden Risiken wirken) sind. Es wird vorausgesetzt, daß T_1^* und T_2^* bedingt auf X unabhängig sind und T_1^* bzw. T_2^* nur die Werte $t_1, \dots, t_{k+1}$ bzw. $s_1, \dots, s_k$ annimmt, wobei $0 < s_1 < t_1 < \dots < s_k < t_k < t_{k+1}$ gelte. ($t_1, \dots, t_k$ repräsentieren bei Anwendungen häufig offene Intervalle. Censorfälle dürfen dann nur an den Intervallgrenzen auftreten. Dies ist zwar strenggenommen meist nicht erfüllt, jedoch ist der Fehler bei nicht zu großen Intervallgrenzen gering.) Beobachtet wird das Minimum T von T_1^* und T_2^* sowie der Indikator $\Delta := I_{(0, \infty)}(T_2^* - T_1^*)$, der angibt, ob ein Failurefall ($\Delta = 1$) oder Censorfall ($\Delta = 0$) vorliegt. Die bedingte Verteilung der Failurezeit T_1^* gegeben $X = x$ werde bis auf einen (ein- oder mehrdimensionalen) Parameter θ als bekannt angenommen:

$$\mathbb{P}_{\theta}\{T_1^* \leq t_i | X = x\} := F_i(\theta, x), \quad i=1, \dots, k,$$

während die Verteilung der Censorzeit T_2^* unspezifiziert bleibt.

Ein wichtiges diskretes Lebenszeitmodell ist das Diskrete - Proportionale - Hazards - Modell (Kalbfleisch, Prentice (80), S.36ff), bei dem die bedingte Verteilung von T_1^* gegeben $X = x$ durch

$$F_i(\theta; x) := 1 - \prod_{j=1}^i (1 - \lambda_j) \exp(\beta'x), \quad i=1, \dots, k$$

mit $\theta := (\lambda_1, \dots, \lambda_k, \beta_1, \dots, \beta_k)'$, $\beta = (\beta_1, \dots, \beta_k)'$ definiert ist.

Sei nun $n \in \mathbb{N}$ die Anzahl der Versuchseinheiten. Beobachtet werden die Zufallsvektoren (T_l, Δ_l, X_l) , $l=1, \dots, n$, welche das oben definierte Lebenszeitmodell erfüllen. Wir nehmen vorläufig an, daß die Vektoren $(T_1, \Delta_1, X_1), \dots, (T_n, \Delta_n, X_n)$ unabhängig sind.

Anstelle von (T_l, Δ_l, X_l) , $l=1, \dots, n$, können wir die Studie auch äquivalent durch die unabhängigen Zufallsvektoren

$(N_{1l}, D_{1l}, \dots, N_{lk}, D_{lk}, X_l)$, $l=1, \dots, n$, beschreiben, wobei $N_{li} := I_{[t_i, \infty)}(T_l)$, $D_{li} := I_{\{t_i\}}(T_l)$, $i=1, \dots, k$, $l=1, \dots, n$, diskrete Zufallsvariablen mit Werten in $\{0, 1\}$ sind, welche die folgenden Bedingungen erfüllen:

- Die bedingte Verteilung von $(N_{1l}, D_{1l}, \dots, N_{lk}, D_{lk})$ gegeben $X_l = x$ hängt nicht von l ab.
- D_{li} ist unter der Bedingung $N_{li} = m$, $X_l = x$ ($m \in \{0, 1\}$) binomialverteilt:

$$\mathcal{L}(D_{li} | N_{li} = m, X_l = x) = \mathcal{B}(m, \rho_i(x, \theta)),$$

wobei $\rho_i(x, \theta) := [F_i(x, \theta) - F_{i-1}(x, \theta)] / [1 - F_{i-1}(x, \theta)]$ mit $F_0(x, \theta) = 0$ gilt.

- D_{li} hängt unter der Bedingung X_l von der Vorgeschichte $(N_{1l}, D_{1l}, \dots, N_{li})$ nur über N_{li} ab:

$$P\{D_{li} = d | N_{1l}, D_{1l}, \dots, N_{li}, X_l\} = P\{D_{li} = d | N_{li}, X_l\} \quad \text{P-f.s.} \\ \forall d \in \mathbb{N}_0.$$

Die Zufallsvariablen N_{li}, D_{li} , $i=1, \dots, k$, $l=1, \dots, n$, lassen sich folgendermaßen in einer $kn \times 2$ -Kontingenztafel zusammenfassen:

D_{11}	$N_{11} - D_{11}$	$\left. \begin{matrix} N_{11} \\ \vdots \\ N_{1k} \end{matrix} \right\} X_1$
D_{1k}	$N_{1k} - D_{1k}$	
D_{n1}	$N_{n1} - D_{n1}$	$\left. \begin{matrix} N_{n1} \\ \vdots \\ N_{nk} \end{matrix} \right\} X_n$
D_{nk}	$N_{nk} - D_{nk}$	

Jede Zeile repräsentiert, bedingt auf die Randsumme und die Kovariable, ein Binomialexperiment. Die asymptotischen ($n \rightarrow \infty$) Eigenschaften dieses Modells werden in Kapitel II behandelt. Die Anzahl der Zellen der Kontingenztafel geht in diesem Fall ins Unendliche (daher der Titel des Kapitels).

Falls der gemeinsame Kovariablenraum der X_n nur aus endlich vielen Punkten - etwa $x_1, \dots, x_s$ - besteht, läßt sich dieses Modell einem allgemeineren, einfacher zu behandelnden Modell unterordnen. Definiere hierzu für jedes $n \in \mathbb{N}$

$$N_{n,s(i-1)+j}^* := \sum_{l=1}^n N_{li} I_{\{x_j\}}(X_l),$$

$$D_{n,s(i-1)+j}^* := \sum_{l=1}^n D_{li} I_{\{x_j\}}(X_l), \quad j=1, \dots, s, \quad i=1, \dots, k.$$

Dann erfüllt für jedes $n \in \mathbb{N}$ $(N_{n1}^*, D_{n1}^*, \dots, N_{nk}^*, D_{nk}^*)$, $k^* := ks$, die Bedingungen

$$- \mathcal{L}(D_{ni}^* | N_{ni}^* = m) = \mathcal{B}(m, \pi_i) \text{ falls } \mathbb{P}\{N_{ni}^* = m\} > 0, \quad i=1, \dots, k^*,$$

$$\text{wobei } \pi_{k(i-1)+j} = \rho_i(x_j, \theta), \quad j=1, \dots, s, \quad i=1, \dots, k.$$

$$- \mathbb{P}\{D_{ni}^* = d | N_{n1}^*, D_{n1}^*, \dots, N_{ni}^*\} = \mathbb{P}\{D_{ni}^* = d | N_{ni}^*\} \quad \mathbb{P}\text{-f.s.}, \quad i=1, \dots, k^* \\ d \in \mathbb{N}_0.$$

Die asymptotischen ($n \rightarrow \infty$) Eigenschaften dieses so definierten Modells werden in Kapitel I behandelt (wobei wir dort das Zeichen "*" fortlassen). Das Modell ist (bei endlichem Kovariablenraum) allgemeiner als das in Kapitel II diskutierte, da sich die Versuchseinheiten nicht unabhängig verhalten müssen. Für jedes $n \in \mathbb{N}$ lassen sich die Zufallsvariablen $N_{n1}^*, D_{n1}^*, \dots, N_{nk^*}^*, D_{nk^*}^*$ wieder ähnlich wie oben in einer Kontingenztafel anordnen:

D_{n1}^*	$N_{n1}^* - D_{n1}^*$	N_{n1}^*
D_{n2}^*	$N_{n2}^* - D_{n2}^*$	N_{n2}^*
$D_{nk^*}^*$	$N_{nk^*}^* - D_{nk^*}^*$	$N_{nk^*}^*$

Hier bleibt für $n \rightarrow \infty$ die Anzahl der Zellen konstant $k^* \geq 2$ (daher der Titel von Kapitel I). In dieses Modell fallen außerdem sowohl Multinomial- als auch Poisson-Stichprobenmodelle (siehe Abschnitt 3). Der Vorteil unseres Ansatzes liegt darin, daß beide Modelle gemeinsam behandelt werden können. Dadurch daß im Modell nur Binomialverteilungen auftreten, sind Teile der Herleitungen einfacher als bei der Behandlung des Multinomialmodells, da keine Nebenbedingungen auftreten.

Der Schwerpunkt dieser Arbeit besteht darin, bei beiden Modellen die Konsistenz und die asymptotische Normalität für eine Klasse von Schätzern herzuleiten. Die in Kapitel I und Kapitel II benutzten Methoden unterscheiden sich stark. Im ersten Kapitel wird im wesentlichen auf Ideen von Rao (73) aufgebaut. Wir schließen dort den für Schärfebetrachtungen notwendigen nichtzentralen Fall mit ein und behandeln außerdem das Testen von Hypothesen. Eine wichtige Rolle spielt in diesem Kapitel das Konzept der Mannigfaltigkeiten, das viele Zusammenhänge geometrisch anschaulicher macht.

Im Kapitel II bringt es keinen Vorteil, sich speziell auf das Kontingenztafelmodell zu beschränken. Wir leiten die asymptotischen Eigenschaften daher für ein allgemeineres Modell her und prüfen anschließend die Regularitätsbedingungen für das Kontingenztafelmodell. Das allgemeine Modell ist auch für sich allein von Interesse, da es andere für Anwendungen wichtige Modelle enthält, z.B. Regressionsmodelle. Die im zweiten Kapitel benutzten Methoden bauen auf die Standardbeweise für Konsistenz (Wald (49) und Witting, Nölle (70)) auf. Daneben stammen einige Ideen aus der Arbeit von Hoadley (71). Die Beweise sind komplizierter und umfangreicher als die im Kapitel I. Um den Rahmen dieser Arbeit nicht zu sprengen, beschränken wir uns deshalb dort auf die Herleitung der Konsistenz und asymptotischen Normalität für den zentralen Fall.

I. KONTINGENZTAFELN MIT ENDLICH VIELEN ZELLEN

2. Modell

Wir betrachten zunächst den finiten Fall und legen folgendes Basismodell zugrunde:

Sei $(\Omega, \mathcal{A}, \mathbb{P}_\pi)$ für jedes $\pi \in (0,1)^k$, $k > 0$, ein Wahrscheinlichkeitsraum und $N_1, \dots, N_k$ und $D_1, \dots, D_k$ diskrete Zufallsvariablen auf $(\Omega, \mathcal{A})$ mit Werten in $\mathbb{N}_0$, welche zwei Eigenschaften erfüllen:

I. Für alle $i=1, \dots, k$ und $\pi = (\pi_1, \dots, \pi_k)' \in (0,1)^k$ ist D_i unter der Bedingung N_i binomialverteilt:

$$\mathcal{L}_{\pi}(D_i | N_i = m) = \mathcal{B}(m, \pi_i) \quad \forall m \in \mathbb{N}_0 \text{ mit } \mathbb{P}_\pi\{N_i = m\} > 0.$$

II. Für alle $i=1, \dots, k$, $\pi \in (0,1)^k$ und $d \in \mathbb{N}_0$ gilt

$$\mathbb{P}_\pi\{D_i = d | N_1, D_1, \dots, N_i\} = \mathbb{P}\{D_i = d | N_i\} \quad \mathbb{P}_\pi\text{-f.s.}$$

Bei Anwendungen ist der wahre Parameter $\pi_0 \in (0,1)^k$ für π - das ist der Parameter, unter dem die vorliegenden Daten erzeugt wurden - im allgemeinen unbekannt, jedoch wird meist vorausgesetzt, daß er in einem bekannten Modellraum $\mathcal{J} \subset (0,1)^k$ liegt. Ziel ist es dann, eine Schätzung $\hat{\pi} \in \mathcal{J}$ des unbekannten Parameters π_0 zu erhalten und seine Güte zu ermitteln. $\mathcal{J}$ ist in der Regel durch eine Parametrisierung

$$\pi(\cdot): \Theta \rightarrow \mathbb{R}^k \quad +)$$

gegeben, wobei $\Theta \subset \mathbb{R}^r$, $0 < r \leq k$, gilt und der offene Kern von Θ nicht leer ist. $\mathcal{J}$ läßt sich dann in der Form

$$\mathcal{J} = \{\pi(\theta) : \theta \in \Theta\}$$

darstellen, wobei $\pi(\cdot) = (\pi_1(\cdot), \dots, \pi_k(\cdot))'$ und $\theta = (\theta_1, \dots, \theta_r)'$ gelten. Wir bezeichnen den wahren Parameter für θ mit θ_0 , so daß $\pi_0 = \pi(\theta_0)$ gilt. Ziel ist es in diesem Fall, eine Schätzung $\hat{\theta} \in \Theta$ für den Parameter θ zu finden. Im folgenden behandeln wir beide Fälle.

+) Da die Gefahr der Verwechslung nicht besteht, benutzen wir den Buchstaben π sowohl für die Bezeichnung der Elemente aus $(0,1)^k$ bzw. $\mathcal{J}$ als auch für die Bezeichnung der Funktion $\Theta \rightarrow \mathbb{R}^k$.

3. Beispiele

3.1 Quantal - Response - Modell (Produkt - Binomialmodell)

$D_1, \dots, D_k$ sind stochastisch unabhängig und besitzen die Verteilung

$$\mathcal{L}_{\pi}(D_i) = \mathcal{B}(n_i, \pi_i), \quad i=1, \dots, k,$$

wobei die Anzahlen $n_i \in \mathbb{N}$ vorgegeben sind. Definiere $N_i, i=1, \dots, k$, als Zufallsvariablen mit Einpunktverteilungen auf $n_i, i=1, \dots, k$, ($N_1 \equiv n_1$). Dann sind die Modellbedingungen I und II erfüllt.

Der Modellraum $\mathcal{Y}$ besitzt häufig die Form

$$(3.1) \quad \mathcal{Y} = \{\pi \in (0,1)^k : \pi_i = F(\theta'x_i), i=1, \dots, k, \theta \in \Theta\},$$

wobei $x_i = (x_{i1}, \dots, x_{ir})' \in \mathbb{R}^r, i=1, \dots, k$, vorgegebene Kovariablen sind und $F(\cdot)$ eine monotone Funktion $\mathbb{R} \rightarrow (0,1)$ ist, z.B. die Normalverteilung (Probitmodell) oder die logistische Verteilung (Logitmodell).

Über das Modell (3.1) liegen eine Reihe von Monographien vor: Finney (71) (nur Probitanalyse), Cox (70), Ashton (72) (beide nur Logitanalyse), Haberman (74, S. 303ff), Linder, Berchtold (76), Schumacher (78). Exakte Beweise finden sich allerdings nur in den Arbeiten von Haberman und Schumacher. Hier sind zusätzlich auch Eigenschaften von Schätzern im finiten Fall aufgeführt.

3.2 Multinomialmodell

$(D_1, \dots, D_{k+1})$ sind multinomialverteilt

$$\mathcal{L}_P(D_1, \dots, D_{k+1}) = \mathcal{M}(n, P),$$

wobei $n \in \mathbb{N}$ vorgegeben ist und $P = (P_1, \dots, P_{k+1})' \in (0,1)^{k+1}$ mit $0 < P_i < 1, i=1, \dots, k+1$, und $P_1 + \dots + P_{k+1} = 1$ ist. Definiere $N_1 \equiv n$, $N_i := N_{i-1} - D_{i-1}, i=2, \dots, k$. Dann erfüllen $N_1, \dots, N_k$ und $D_1, \dots, D_k$ die Modellbedingungen I und II, wobei

$$\pi_i := P_i / \sum_{j=1}^{k+1} P_j, \quad i=1, \dots, k,$$

gelten.

Selbstverständlich erfüllt damit auch das Produkt - Multinomialmodell (Zusammenschluß mehrerer unabhängiger Multinomialexperimente) diese Bedingungen.

Die asymptotischen Eigenschaften von Maximum - Likelihood - Schätzern im Multinomialmodell werden im Buch von Rao (73, S. 355ff) hergeleitet. Der spezielle Fall des log - linearen Modells wird im Buch von Haberman (74) ausführlich behandelt, wobei dort zusätzlich Poisson - (siehe Beispiel 3.3) und Bedingte - Poisson - Stichprobenmodelle eingeschlossen werden.

3.3 Poissonmodell

$(D_1, \dots, D_{k+1})$ besitzt die Zähldichte

$$f_{P_1, \dots, P_{k+1}}(d_1, \dots, d_{k+1}) = \prod_{i=1}^{k+1} \frac{(a \cdot P_i)^{d_i}}{d_i!} e^{-a \cdot P_i}, \quad a > 0, \\ d_1, \dots, d_{k+1} \in \mathbb{N}_0^{k+1},$$

wobei $0 < P_i < 1$, $i=1, \dots, k+1$, und $P_1 + \dots + P_{k+1} = 1$ gelten. Definiere $N_i := D_1 + \dots + D_{k+1}$, $i=1, \dots, k$. Dann erfüllen $N_1, \dots, N_k$ und $D_1, \dots, D_k$ die Modellbedingungen I und II, wobei

$$\pi_i := P_i / \sum_{j=1}^{k+1} P_j, \quad i=1, \dots, k,$$

gelten. Da die bedingte Wahrscheinlichkeit von D_i gegeben N_i nicht von a abhängt, kann bei diesem Modell natürlich keine Aussage über diesen Parameter gemacht werden.

3.4 Diskretes Lebenszeitmodell

Bei diesem in der Einleitung diskutierten Modell haben die Zufallsvariablen folgende Bedeutung:

$N_{s(i-1)+j}$: Anzahl der Versuchseinheiten, die dem Kovariablenwert x_j zugeordnet sind und im Zeitpunkt t_i dem Risiko ausgesetzt sind.

$D_{s(i-1)+j}$: Anzahl der Versuchseinheiten, die dem Kovariablenwert x_j zugeordnet sind und im Zeitpunkt t_i aufgrund des Risikos ausfallen.

Hierbei ist s die Anzahl der verschiedenen Kovariablenwerte. Die Modellbedingungen sind auch unter sehr allgemeinem Censormechanismus noch erfüllt.

Im speziellen Diskreten - Proportionalem - Hazards - Modell (siehe Einleitung) gelten

$$\pi_{s(i-1)+j} = F(\gamma_i + \beta'x_j), \quad i=1, \dots, k, \quad j=1, \dots, s,$$

wobei $F(u) := 1 - \exp(-e^u)$ die komplementäre log-log - Verteilung und $\gamma_i := \log[-\log(1-\lambda_i)]$, $i=1, \dots, k$, sind.

4. Schätzmethoden

Um den Schätzer für π_o bzw. θ_o zu ermitteln, gehen wir von einer Funktion $\mathcal{D}(\cdot): [0,1]^k \times (0,1)^k \times [0,\infty)^k \rightarrow [0,\infty)$ aus, welche die Form

$$(4.1) \quad \mathcal{D}(p, \pi, q) = \sum_{i=1}^k q_i h_i(p_i, \pi_i)$$

besitzt, wobei $p = (p_1, \dots, p_k)'$, $\pi = (\pi_1, \dots, \pi_k)'$, $q = (q_1, \dots, q_k)'$ sind und $h_i(\cdot)$, $i=1, \dots, k$, sogenannte Abstandsfunktionen $[0,1] \times (0,1) \rightarrow [0,\infty)$ sind.

Im folgenden seien $N := (N_1, \dots, N_k)'$ und $P := (P_1, \dots, P_k)'$ mit

$$(4.2) \quad P_i := \begin{cases} \frac{D_i}{N_i}, & \text{falls } N_i > 0, D_i \leq N_i \\ 0, & \text{sonst} \end{cases}, \quad i=1, \dots, k.$$

Der Schätzer $\hat{\pi}$ bzw. $\hat{\theta}$ bzgl. $\mathcal{D}(\cdot)$ unter $\mathcal{J}$ bzw. Θ sei irgendeine (nicht notwendig meßbare) Variable $\Omega \rightarrow \mathcal{J}$ bzw. $\Omega \rightarrow \Theta$, welche die Gleichung

$$(4.3) \quad \mathcal{D}(P, \hat{\pi}, N) = \inf_{\pi \in \mathcal{J}} \mathcal{D}(P, \pi, N)$$

bzw.

$$(4.4) \quad \mathcal{D}(P, \pi(\hat{\theta}), N) = \inf_{\theta \in \Theta} \mathcal{D}(P, \pi(\theta), N)$$

erfüllt, sofern das Infimum für ein $\pi \in \mathcal{J}$ bzw. $\theta \in \Theta$ angenommen wird.

Im folgenden werden vier gebräuchliche Schätzer vorgestellt. Weitere sinnvolle Schätzer findet man z.B. in Rao (73, S.352).

Maximum - Likelihood - Schätzer :

Diesem Schätzer kommt aus theoretischen Gründen die größte Bedeutung zu (näheres siehe z.B. Rao (73, S.348) und wird am häufigsten angewandt. Die volle Likelihoodfunktion läßt sich

bei unserem Modell nicht aufstellen, da nur Teilinformation über die Verteilung von $(N_1, D_1, \dots, N_k, D_k)$ zur Verfügung steht. Man kann jedoch eine partielle Likelihoodfunktion (Cox(75)) aufstellen.

Seien hierzu $N^i := (N_1, \dots, N_i)$, $D^i := (D_1, \dots, D_i)$, $i=1, \dots, k$, und $n^i := (n_1, \dots, n_i)$, $d_i := (d_1, \dots, d_i)$, $i=1, \dots, k$, zugehörige Realisierungen. Seien weiter $f_{N_i | N^{i-1}, D^{i-1}}(\cdot; \pi)$ und $f_{D_i | N^i, D^{i-1}}(\cdot; \pi)$ die bedingten Dichten von N_i und D_i (bzgl. des Zählmaßes) unter π jeweils bedingt auf die gesamte "Vorgeschichte". Dann läßt sich die (volle) Likelihoodfunktion schreiben als:

$$\prod_{i=1}^k f_{N_i | N^{i-1}, D^{i-1}}(n_i | n^{i-1}, d^{i-1}; \pi) \cdot f_{D_i | N^i, D^{i-1}}(d_i | n^i, d^{i-1}; \pi).$$

Die Dichten $f_{N_i | N^{i-1}, D^{i-1}}(\cdot; \pi)$ sind in unserem Modell nicht näher spezifiziert, den übrigen Teil der Likelihoodfunktion - eine partielle Likelihoodfunktion - können wir jedoch angeben:

$$(4.5) \quad \prod_{i=1}^k f_{D_i | N^i, D^{i-1}}(d_i | n^i, d^{i-1}; \pi) = \prod_{i=1}^k \pi_i^{d_i} (1-\pi_i)^{n_i-d_i}.$$

Maximierung dieser partiellen Likelihoodfunktion ist äquivalent zu der Minimierung von

$$\sum_{i=1}^k n_i \left[p_i \log \frac{p_i}{\pi_i} + (1-p_i) \log \frac{1-p_i}{1-\pi_i} \right],$$

wenn $p_i := \begin{cases} d_i/n_i & , \text{ falls } n_i > 0 \\ 0 & \text{sonst} \end{cases}$ und $0 \log 0 := 0$ gesetzt wird.

Wir definieren daher die Abstandsfunktion $h_i(\cdot)$ durch

$$(4.6) \quad h_i(a, b) := a \log \frac{a}{b} + (1-a) \log \frac{1-a}{1-b}, \quad i=1, \dots, k.$$

In Kapitel II ist es bequemer, anstelle (4.6) die Abstandsfunktion

$$(4.7) \quad h_i(a, b) := -a \log b - (1-a) \log (1-b), \quad i=1, \dots, k,$$

zu nehmen. Beide Abstandsfunktionen liefern offensichtlich denselben Schätzer, da sie sich (bei fester Realisierung) nur durch eine Konstante unterscheiden.

Dieser Schätzer $\hat{\pi}$ bzw. $\hat{\theta}$ ist natürlich im allgemeinen nicht gleich dem vollen Maximum-Likelihood-Schätzer, da die mögliche Information, die die bedingten Dichten $f_{N_i | N^{i-1}, D^{i-1}}(\cdot)$, $i=1, \dots, k$, für π liefern, nicht berücksichtigt werden. Hängen diese Dichten jedoch nicht von π ab, so ist (4.5) proportional zu der vollen Likelihood und $\hat{\pi}$ bzw. $\hat{\theta}$ daher der volle Maximum-Likelihood-Schätzer. Dies ist z.B. bei den oben erwähnten Quantal-Response Modell, Multinomialmodell und Poissonmodell der Fall, im allgemeinen jedoch nicht beim Diskreten-Proportionalem-Hazards-Modell (falls nämlich der Censormechanismus vom Regressionsparameter β abhängt).

Minimum - χ^2 - Schätzer :

Die Abstandsfunktionen sind durch

$$(4.8) \quad h_i(a, b) := \frac{(a - b)^2}{b(1 - b)}, \quad i=1, \dots, k,$$

definiert. Motiviert wird dieser Schätzer dadurch, daß wegen Modellannahme I für alle $\frac{1}{n} \cdot B(n, b)$ -verteilten ($n > 0$, $0 < b < 1$) Zufallsvariablen P

$$\text{Var}(P) = \frac{b(1 - b)}{n}$$

gilt.

Verallgemeinerter modifizierter Minimum - χ^2 - Schätzer :

Die Abstandsfunktionen $h_i(\cdot)$ sind derart auf $[0, 1] \times (0, 1)$ definiert, daß

$$(4.9) \quad h_i(a, b) := \frac{[w(a) - w(b)]^2}{a(1 - a)} \cdot \left(\frac{1}{w'(a)} \right)^2 \quad \forall a \in (0, 1), \quad b \in (0, 1), \\ i=1, \dots, k,$$

gilt, wobei $w(\cdot)$ eine differenzierbare Funktion $(0,1) \rightarrow \mathbb{R}$ ist und $w'(u) := \frac{dw}{du}(u) > 0$ für alle $u \in (0,1)$ gilt. Motiviert wird dieser Schätzer dadurch, daß jede Folge $\{P_n\}$ von $\frac{1}{n} \cdot \mathcal{K}(n,a)$ -verteilten Zufallsvariablen ($0 < a < 1$) die Beziehung

$$\text{Var}\{w(P_n)\} = [w'(a)]^2 \cdot \frac{a(1-a)}{n} + o\left(\frac{1}{n}\right) \quad (n \rightarrow \infty)$$

erfüllt und a konsistent durch P_n geschätzt werden kann.

Häufig läßt sich $w(\cdot)$ derart wählen, daß $\{(w(\pi_1), \dots, w(\pi_k)) : \pi \in \mathcal{P}\}$ eine lineare Mannigfaltigkeit im $\mathbb{R}^k$ ist. In diesem Fall läßt sich der Schätzer $\hat{\pi}$ explizit angeben.

Falls $w(\cdot)$ die Identität ist, heißt dieser Schätzer "modifizierter Minimum- χ^2 -Schätzer".

Einfacher Schätzer:

Die Abstandsfunktionen sind durch

$$(4.10) \quad h_i(a,b) := c_i(a-b)^2, \quad i=1, \dots, k,$$

definiert, wobei $c_i > 0$, $i=1, \dots, k$, vorgegebene Gewichte sind. Diese Methode ist sinnvoll, wenn die Varianzen aufgrund von Versuchen annähernd bekannt sind.

Man beachte, daß beim Multinomialmodell (Abschnitt 3.2) die durch (4.8) und (4.9) definierten Minimum- χ^2 -Schätzer nicht mit den üblichen Minimum- χ^2 -Schätzern übereinstimmen.

5. Konsistenz

Um das asymptotische Verhalten des Schätzers $\hat{\pi}$ bzw. $\hat{\theta}$ zu untersuchen, betrachten wir die Folgen $\{N_n\}$ und $\{D_n\}$ von Zufallsvektoren auf $(\Omega, \mathcal{O})$, wobei $N_n := (N_{n1}, \dots, N_{nk})'$ und $D_n := (D_{n1}, \dots, D_{nk})'$, $n \in \mathbb{N}$, sind und (N_n, D_n) für jedes $n \in \mathbb{N}$ die Modellannahme I erfüllt. (II wird für die Konsistenz nicht benötigt. P_n und der Schätzer $\hat{\pi}_n$ bzw. $\hat{\theta}_n$ seien analog P , $\hat{\pi}$ bzw. $\hat{\theta}$ definiert, wenn wir in Abschnitt 4 N und D durch N_n und D_n ersetzen. Den wahren Parameter für π bezeichnen wir mit π_{on} ($\pi_{on} \in (0,1)^k$); er darf von n abhängen und muß nicht in $\mathcal{S}$ liegen, soll aber gegen einen Punkt $\pi_0 \in \mathcal{S}$ konvergieren.

Ziel dieses Abschnitts ist es zu zeigen, daß $\hat{\pi}_n$ bzw. $\hat{\theta}_n$ unter der Folge von Wahrscheinlichkeitsmaßen $\{\mathbb{P}_{\pi_{on}}\}$ konsistent für π_0 bzw. θ_0 , definiert durch $\pi_0 = \pi(\theta_0)$, ist. Im folgenden schreiben wir anstelle des Wahrscheinlichkeitsmaßes $\mathbb{P}_{\pi_{on}}$ kurz $\mathbb{P}_n$ und entsprechend für die Verteilung, den Erwartungswert und die Varianz unter π_{on} kurz $\mathcal{L}_n$, E_n und Var_n .

Eine sinnvolle Forderung für die Konsistenz ist sicherlich, daß die Anzahl der Versuchseinheiten nach Wahrscheinlichkeit ins Unendliche wächst:

Annahme A:

$$S_n := \sum_{i=1}^k N_{ni} \xrightarrow{\mathbb{P}_n} \infty \quad (n \rightarrow \infty).$$

Außerdem fordern wir, daß der relative Anteil jeder Gruppe an der Gesamtanzahl S_n nach Wahrscheinlichkeit gegen eine positive Konstante konvergiert:

Annahme B:

Es existiert ein positiver Vektor $\eta = (\eta_1, \dots, \eta_k)' \in (0,1)^k$, so daß

$$N_n / S_n \xrightarrow{\mathbb{P}_n} \eta \quad (n \rightarrow \infty)$$

gilt.

Lemma 5.1

Erfüllt für jedes $n \in \mathbb{N}$ (N_n, D_n) die Modellannahme I sowie die Annahmen A und B, so gilt

$$P_n \xrightarrow{\mathbb{P}_n} \pi_0.$$

Beweis:

Sei $i \in \{1, \dots, k\}$ und $\varepsilon > 0$ fest. Definiere die Mengen

$$A_n = \left\{ \left| \frac{D_{ni}}{N_{ni}} - \pi_{oni} \right| \leq \varepsilon \right\} \quad \forall n \in \mathbb{N}.$$

Dann gilt für beliebiges $m_0 > 0$

$$(5.1) \quad \mathbb{P}_n(A_n) = \int_{\{N_{ni} > m_0\}} \mathbb{P}_n(A_n | N_{ni}) d\mathbb{P}_n.$$

Nach der Tschebyscheff-Ungleichung folgt aus Modellannahme I für $N_{ni} > 0$

$$\begin{aligned} \mathbb{P}_n(A_n | N_{ni}) &\geq 1 - \frac{\pi_{oni}(1 - \pi_{oni})}{N_{ni} \varepsilon^2} \\ &\geq 1 - \frac{1}{N_{ni} \varepsilon^2} \quad \mathbb{P}_n\text{-f.s.} \end{aligned}$$

Dies in (5.1) eingesetzt liefert

$$\mathbb{P}_n(A_n) \geq \left(1 - \frac{1}{m_0 \varepsilon^2}\right) \mathbb{P}_n\{N_{ni} > m_0\}.$$

Da wegen Annahmen A und B $\mathbb{P}_n\{N_{ni} > m_0\} \rightarrow 1$ gilt und m_0 beliebig wählbar ist, folgt somit

$$\mathbb{P}_n(A_n) \rightarrow 1.$$

Da $\varepsilon > 0$ beliebig wählbar ist, erhalten wir hieraus

$$\frac{D_{ni}}{N_{ni}} - \pi_{oni} \xrightarrow{\mathbb{P}_n} 0.$$

Wegen $\pi_{oni} \rightarrow \pi_{oi}$ ist damit die Behauptung bewiesen.

An $\mathcal{P}$ und die Abstandsfunktionen $h_i(\cdot)$, $i=1, \dots, k$, stellen wir die folgenden Regularitätsvoraussetzungen:

K1: $\mathcal{P}$ ist lokalkompakt (d.h. für alle $\pi \in \mathcal{P}$ existiert eine abgeschlossene Umgebung K um π , so daß $K \cap \mathcal{P}$ kompakt ist).

K2: Die Funktionen $h_i(\cdot)$, $i=1, \dots, k$, sind stetig auf $(0,1) \times (0,1)$.

K3: Für alle $a \in (0,1)$ und alle $\varepsilon > 0$ existiert ein $\delta > 0$, so daß für alle $b \in (0,1)$ mit $|a-b| > \varepsilon$ die Ungleichungen

$$h_i(a,b) - h_i(a,a) > \delta, \quad i=1, \dots, k,$$

erfüllt sind.

K3 stellt sicher, daß a der Punkt ist, an dem die Funktionen $h_i(a, \cdot)$ ihr Minimum annehmen, und daß diese Funktionen außerhalb einer beliebig kleinen Umgebung um a dem Wert $h_i(a,a)$ nicht beliebig nahe kommen (strenge Identifizierbarkeit von a durch $h_i(a, \cdot)$).

K4: Für alle $a \in (0,1)$, alle $M > 0$ und alle $\varepsilon > 0$ existiert ein $\delta > 0$, so daß für alle $b \in (0,1)$ mit $|a-b| < \delta$ und alle $\rho \in (0,1)$ die Ungleichungen

$$h_i(b, \rho) > \min\{M, h_i(a, \rho)\} - \varepsilon, \quad i=1, \dots, k,$$

gelten.

K4 drückt aus, daß für alle $M > 0$ die Funktionen $\min\{M, h_i(\cdot, \rho)\}$ gleichmäßig in ρ unten halbstetige Funktionen sind.

Bei den im Abschnitt 4 definierten Abstandsfunktionen sind die Regularitätsbedingungen K3 und K4 erfüllt, wie wir im folgenden zeigen werden.

Maximum-Likelihood-Schätzer:

Die Gültigkeit von K3 folgt sofort wegen

$$D_2 h_1(a,b) = \frac{b-a}{b(1-b)} \begin{cases} < 0, & \text{falls } b < a \\ = 0, & \text{" } b = a, \quad \forall a, b \in (0,1). \\ > 0, & \text{" } b > a \end{cases}$$

Seien weiter $\varepsilon > 0$, $a \in (0,1)$, $M > 0$ und $0 < \delta_0 < \min\{a, 1-a\}$ beliebig aber fest. Sei außerdem

$$A := \{\rho \in (0,1): a \log a + (1-a) \log(1-a) > M + (a-\delta_0) \log \rho + (1-a-\delta_0) \log(1-\rho)\}.$$

Dann existiert ein $0 < \delta < \delta_0$, so daß für alle $b \in (0,1)$ mit $|a-b| < \delta$ und alle $\rho \in A$

$$\begin{aligned} b \log b + (1-b) \log(1-b) &> a \log a + (1-a) \log(1-a) - \varepsilon \\ &> M - \varepsilon + (a-\delta_0) \log \rho + (1-a-\delta_0) \log(1-\rho) \\ &> M - \varepsilon + b \log \rho + (1-b) \log(1-\rho) \end{aligned}$$

und daher

$$h_1(b, \rho) > M - \varepsilon$$

gilt. Für $\rho \in (0,1)-A$ sind $\log \rho$ und $\log(1-\rho)$ gleichmäßig in ρ nach unten beschränkt, etwa durch $-K < 0$. Außerdem existiert ein $\tilde{\delta} < 0$, so daß für alle $b \in (0,1)$ mit $|a-b| < \tilde{\delta}$

$$|a \log a + (1-a) \log(1-a) - b \log b - (1-b) \log(1-b)| \leq \frac{\varepsilon}{2}$$

gilt. Damit folgt für alle $b \in (0,1)$ mit $|a-b| < \delta := \min\{\tilde{\delta}, \varepsilon/4K\}$

$$\begin{aligned} h_1(b, \rho) &\geq h_1(a, \rho) - \frac{\varepsilon}{2} - |a-b| \cdot |\log \rho + \log(1-\rho)| \\ &\geq h_1(a, \rho) - \frac{\varepsilon}{2} - \delta 2K \\ &\geq h_1(a, \rho) - \varepsilon. \end{aligned}$$

Damit ist K_4 gezeigt.

Minimum - x^2 - Schätzer :

K_3 folgt sofort wegen

$$h_1(a, b) \geq (a-b)^2 \quad \forall a, b \in (0,1).$$

Seien weiter $\varepsilon > 0$, $a \in (0,1)$, $M > 0$ und $0 < \delta_0 < \min\{a, 1-a\}$ beliebig. Wir betrachten zunächst die Menge

$$A := \{\rho \in (0,1): |a-\rho| - \delta_0 > \sqrt{M \rho(1-\rho)}\}.$$

Für alle $\rho \in A$ und alle $b \in (0,1)$ mit $|a-b| < \delta_0$ gilt dann

$$|b-\rho| > |a-\rho| - \delta_0 > \sqrt{M\rho(1-\rho)}$$

und daher

$$h_1(b, \rho) = \frac{(b-\rho)^2}{\rho(1-\rho)} > M.$$

Für $\rho \in (0,1)-A$ ist $\rho(1-\rho)$ gleichmäßig nach unten beschränkt, etwa durch $K > 0$. In diesem Fall folgt für alle $b \in (0,1)$ mit $|a-b| < \delta := \varepsilon K/2$

$$\begin{aligned} h_1(b, \rho) &= \frac{(b-\rho)^2}{\rho(1-\rho)} \\ &> \frac{(a-\rho)^2}{\rho(1-\rho)} - \left| \frac{(a-\rho)^2 - (b-\rho)^2}{\rho(1-\rho)} \right| \\ &> h_1(a, \rho) - \frac{|a-b| \cdot |a+b-2\rho|}{K} \\ &> h_1(a, \rho) - \delta \frac{2}{K} > h_1(a, \rho) - \varepsilon. \end{aligned}$$

Damit ist $K4$ bewiesen.

Verallgemeinerter modifizierter Minimum- x^2 -Schätzer :

$K3$ gilt wegen

$$\begin{aligned} D_2 h_1(a, b) &= 2 \frac{w(b) - w(a)}{a(1-a)} \frac{w'(b)}{[w'(a)]^2} \\ &\begin{cases} < 0, & \text{falls } b < a \\ = 0, & \text{" } b = a \\ > 0, & \text{" } b > a \end{cases}, \quad a, b \in (0,1). \end{aligned}$$

Seien weiter $\varepsilon > 0$, $a \in (0,1)$, $M > 0$ und $\delta_0 > 0$. Außerdem sei

$$A := \left\{ \rho \in (0,1) : |w(a) - w(\rho)| - \delta_0 > \sqrt{\frac{M}{f(a)}} \right\},$$

wobei $f(a) := \frac{1}{a(1-a)[w'(a)]^2}$ gelte. Dann folgt für alle $b \in (0,1)$

mit $|w(a) - w(b)| < \delta_0$ und alle $\rho \in A$

$$h_1(b, \rho) = f(b) (w(b) - w(\rho))^2 > M \frac{f(b)}{f(a)}.$$

Wegen der Stetigkeit von $f(\cdot)$ und $w(\cdot)$ erhalten wir somit die Existenz eines $\delta > 0$, so daß aus $|a-b| < \delta$

$$h_1(b, \rho) > M - \epsilon$$

folgt. Für $\rho \in (0, 1) - A$ ist $|w(a) - w(\rho)|$ gleichmäßig in ρ nach oben beschränkt, etwa durch $K > 0$. Wir erhalten somit für diesen Fall

$$\begin{aligned} h_1(b, \rho) &= f(b) (w(b) - w(\rho))^2 \\ &\geq f(a) (w(a) - w(\rho))^2 - |f(a) - f(b)| (w(a) - w(\rho))^2 - \\ &\quad - f(b) |w(a) - w(b)| (w(a) + w(b) - 2w(\rho))^2 \\ &> h_1(a, \rho) - |f(a) - f(b)| K^2 - \\ &\quad - f(b) |w(a) - w(b)| (|w(a) - w(b)| + 2K). \end{aligned}$$

Wegen der Stetigkeit von $f(\cdot)$ und $w(\cdot)$ gibt es daher ein $\delta > 0$, so daß für alle $\rho \in (0, 1) - A$ und $b \in (0, 1)$ mit $|a-b| < \delta$

$$h_1(b, \rho) > h_1(a, \rho) - \epsilon$$

folgt. Damit ist K_4 gezeigt.

Einfacher Schätzer :

K_3 erhält man sofort. K_4 folgt mit der Tatsache, daß für alle $a, b, \rho \in (0, 1)$

$$h_i(b, \rho) = c_i(b - \rho)^2 \geq c_i(a - \rho)^2 - 2c_i|a - b|, \quad i=1, \dots, k,$$

gilt.

Lemma 5.2

Sei $\eta = (\eta_1, \dots, \eta_k)' \in (0, 1)^k$, Aus K3 folgt:

(i) Für alle $\varepsilon > 0$ existiert ein $\delta > 0$ derart, daß für alle $\pi \in \mathcal{P}$ mit $\|\pi - \pi_0\| > \varepsilon$

$$\mathcal{D}(\pi_0, \pi, \eta) - \mathcal{D}(\pi_0, \pi_0, \eta) > \delta$$

gilt.

Außerdem folgt aus K4:

(ii) Für alle $M > 0$ und $\varepsilon > 0$ existiert ein $\delta > 0$ derart, daß für alle $\pi \in \mathcal{P}$, $p \in (0, 1)^k$, $q \in (0, 1)^k$ die Implikation

$$\mathcal{D}(\pi_0, \pi, \eta) > M, \quad \|\pi - \pi_0\| < \delta, \quad \|q - \eta\| < \delta \implies \mathcal{D}(p, \pi, q) > M - \varepsilon$$

gilt.

Beweis:

Seien $\varepsilon > 0$, $\eta = (\eta_1, \dots, \eta_k)' \in (0, 1)^k$ beliebig und

$$A_i := \{\pi \in \mathcal{P} : |\pi_i - \pi_{0i}| > \frac{\varepsilon}{k}\}, \quad i=1, \dots, k.$$

Dann existiert wegen K3 für jedes $i=1, \dots, k$ ein $\delta_i > 0$, so daß aus $\pi \in A_i$

$$h_i(\pi_{0i}, \pi_i) - h_i(\pi_{0i}, \pi_{0i}) > \delta_i$$

und damit

$$\mathcal{D}(\pi_0, \pi, \eta) - \mathcal{D}(\pi_0, \pi_0, \eta) > \delta_i \eta_i$$

folgt. Wegen

$$\pi \in \mathcal{P}, \quad \|\pi - \pi_0\| > \varepsilon \implies \pi \in A_i \quad \text{für ein } i \in \{1, \dots, k\},$$

erhalten wir (i), indem wir $\delta := \min\{\delta_1 \eta_1, \dots, \delta_k \eta_k\}$ setzen.

Gelte zusätzlich $M > 0$. Dann existiert wegen K4 ein $\delta > 0$, so daß für alle $p_i \in (0, 1)$, $q_i \in (0, 1)$ mit $|p_i - \pi_{0i}| < \delta$, $|q_i - \eta_i| < \delta$, $i=1, \dots, k$, und $\pi \in \mathcal{P}$

$$\begin{aligned}
 (5.2) \quad q_i h_i(p_i, \pi_i) &> q_i \min\left\{\frac{M}{\eta_i}, h_i(\pi_{oi}, \pi_i)\right\} - \frac{\xi}{2k} \\
 &> \min\{M, \eta_i h_i(\pi_{oi}, \pi_i)\} - \frac{\xi}{k}, \quad i=1, \dots, k,
 \end{aligned}$$

folgt. Hierbei gilt die zweite Ungleichung, da für genügend kleines $\delta > 0$ im Fall $M \leq \eta_i h_i(\pi_{oi}, \pi_i)$

$$q_i \min\left\{\frac{M}{\eta_i}, h_i(\pi_{oi}, \pi_i)\right\} = q_i \frac{M}{\eta_i} > M - \frac{\xi}{2k}$$

und im Fall $M > \eta_i h_i(\pi_{oi}, \pi_i)$

$$\begin{aligned}
 q_i \min\left\{\frac{M}{\eta_i}, h_i(\pi_{oi}, \pi_i)\right\} &\geq \eta_i h_i(\pi_{oi}, \pi_i) - |q_i - \eta_i| h_i(\pi_{oi}, \pi_i) \\
 &> \eta_i h_i(\pi_{oi}, \pi_i) - \delta \frac{M}{\eta_i} \\
 &> \eta_i h_i(\pi_{oi}, \pi_i) - \frac{\xi}{2k}
 \end{aligned}$$

gelten. Da aus (5.2) die Ungleichung

$$\sum_{i=1}^k q_i h_i(p_i, \pi_i) > \min\left\{M, \sum_{i=1}^k \eta_i h_i(\pi_{oi}, \pi_i)\right\} - \xi$$

folgt, ist (ii) gezeigt. □

Mit Lemma 5.1 und Lemma 5.2 läßt sich die (schwache) Konsistenz der Folge $\{\hat{\pi}_n\}$ nun einfach zeigen. Da $\hat{\pi}_n$ und auch $\hat{\theta}_n$ im allgemeinen nicht meßbar sind, benutzen wir im folgenden eine allgemeinere Definition für die Konvergenz nach Wahrscheinlichkeit ($\xrightarrow{\mathbb{P}}_n$) und nach Verteilung ($\xrightarrow{\mathcal{L}}_n$), die den Fall nicht meßbarer Variablen einschließt (siehe hierzu Anhang B).

Satz 5.3

Sind die Modellvoraussetzung I, die Annahmen A und B über die Asymptotik sowie die Regularitätsannahmen K1 - K4 erfüllt, so gelten

$$\mathbb{P}_{n*}\{\mathcal{D}(P_n, \hat{\pi}_n, N_n) = \inf_{\pi \in \mathcal{Y}} \mathcal{D}(P_n, \pi, N_n)\} \longrightarrow 1 \quad +)$$

und

$$\hat{\pi}_n \xrightarrow{\mathbb{P}_n} \pi_0.$$

Beweis:

Sei $\varepsilon > 0$ beliebig und sei

$$\bar{A}_\varepsilon := \{\pi \in \mathcal{Y} : \|\pi - \pi_0\| \leq \varepsilon\}.$$

Wegen K3 und Lemma 5.2(i) existiert ein $\delta > 0$, so daß für alle $\pi \in \mathcal{Y} - \bar{A}_\varepsilon$ die Ungleichung

$$\mathcal{D}(\pi_0, \pi, \eta) - \mathcal{D}(\pi_0, \pi_0, \eta) > 3\delta$$

gilt. Hieraus folgt mit K4 und Lemma 5.2(ii) die Existenz eines $\gamma > 0$, so daß für alle $p \in (0,1)^k$ und $q \in (0,1)^k$ mit $\|q - \eta\| < \gamma$, $\|p - \pi_0\| < \gamma$ und alle $\pi \in \mathcal{Y} - \bar{A}_\varepsilon$ die Ungleichung

$$(5.3) \quad \mathcal{D}(p, \pi, q) > 2\delta + \mathcal{D}(\pi_0, \pi_0, \eta)$$

gilt. Bei genügend kleinem $\gamma > 0$ gilt außerdem wegen der Stetigkeit von $\mathcal{D}(\cdot)$ auf $(0,1)^k \times (0,1)^k \times \mathbb{R}^k$ (K2)

$$\mathcal{D}(p, \pi_0, q) > \delta + \mathcal{D}(\pi_0, \pi_0, \eta)$$

und daher wegen (5.3)

$$\mathcal{D}(p, \pi, q) > \delta + \mathcal{D}(p, \pi_0, q).$$

+) $\mathbb{P}_{n*}$ ist das im Anhang B definierte innere Wahrscheinlichkeitsmaß zu $\mathbb{P}_n$.

Damit haben wir bis jetzt bewiesen: für alle $\varepsilon > 0$ gibt es eine Umgebung U um (π_0, η) , so daß für alle $(p, q) \in U$

$$\inf_{\pi \in \mathcal{P} - \bar{A}_\varepsilon} \mathcal{D}(p, \pi, q) > \mathcal{D}(p, \pi_0, q) \geq \inf_{\pi \in \mathcal{P}} \mathcal{D}(p, \pi, q)$$

gilt. Wir erhalten somit wegen der Stetigkeit von $\mathcal{D}(p, \cdot, q)$

$$\begin{aligned} (5.4) \quad & \{(P_n, N_n / S_n) \in U\} \\ & \subset \left\{ \inf_{\pi \in \mathcal{P} - \bar{A}_\varepsilon} \mathcal{D}(P_n, \pi, N_n) > \inf_{\pi \in \bar{A}_\varepsilon} \mathcal{D}(P_n, \pi, N_n) \right\} \\ & \subset \left\{ \inf_{\pi \in \mathcal{P} - \bar{A}_\varepsilon} \mathcal{D}(P_n, \pi, N_n) > \mathcal{D}(P_n, \hat{\pi}_n, N_n) = \inf_{\pi \in \bar{A}_\varepsilon} \mathcal{D}(P_n, \pi, N_n) \right\} \\ & \subset \{\hat{\pi}_n \in \bar{A}_\varepsilon\}. \end{aligned}$$

Wegen $N_n / S_n \xrightarrow{\mathbb{P}_n} \eta$ (Annahme B) und $P_n \xrightarrow{\mathbb{P}_n} \pi_0$ (Lemma 5.1) folgen aus (5.4) die erste Behauptung und

$$\mathbb{P}_n * \{\hat{\pi}_n \in \bar{A}_\varepsilon\} \rightarrow 1.$$

Da $\varepsilon > 0$ beliebig ist, ist damit auch die zweite Behauptung bewiesen. □

An dem Beweis erkennt man, daß für die Konsistenz nicht erforderlich ist, daß $\mathcal{D}(\cdot)$ die spezielle Form (4.1) besitzt. Benötigt werden dann anstelle von K2-K4 die Stetigkeit von $\mathcal{D}(\cdot)$ sowie die Folgerungen (i) und (ii) in Lemma 5.2.

Der Satz läßt sich leicht auf den Fall verallgemeinern, daß π_0 nicht im Modellraum $\mathcal{P}$ liegt. $\hat{\pi}_n$ konvergiert dann nach Wahrscheinlichkeit gegen einen Punkt $\pi^* \in \mathcal{P}$, der durch

$$\mathcal{D}(\pi_0, \pi^*, \eta) = \inf_{\pi \in \mathcal{P}} \mathcal{D}(\pi_0, \pi, \eta)$$

definiert ist, also bzgl. $\mathcal{D}(\cdot)$ am nächsten zu π_0 liegt. Die Regularitätsannahme K3 bzw. die Folgerung (i) in Lemma 5.1 muß dann ersetzt werden durch:

Es existiert ein $\pi^* \in \mathcal{Y}$ derart, daß es für alle $\varepsilon > 0$ ein $\delta > 0$ gibt mit

$$\pi \in \mathcal{Y}, \|\pi - \pi^*\| > \varepsilon \implies \mathcal{D}(\pi_0, \pi, \eta) - \mathcal{D}(\pi_0, \pi^*, \eta) > \delta.$$

Der Beweis der Konsistenz $\hat{\pi}_n \xrightarrow{\mathbb{P}_n} \pi^*$ verläuft ähnlich dem Beweis zu Satz 5.3. Allerdings ist die Regularitätsannahme für einzelne Beispiele sehr schwierig nachzuprüfen, da sie über π^* mit dem Modellraum $\mathcal{Y}$ zusammenhängt.

Ist $\mathcal{Y}$ durch eine Parametrisierung

$$\mathcal{Y} := \{\pi(\theta) : \theta \in \Theta\}, \quad \Theta \subset \mathbb{R}^r, \quad 0 \leq r \leq k,$$

gegeben und wollen wir die (schwache) Konvergenz der Folge $\{\hat{\theta}_n\}$ zeigen, wobei $\hat{\theta}_n$ entsprechend (4.3) definiert ist, so benötigen wir weitere Regularitätsannahmen:

K5: Θ ist lokalkompakt.

Die Parametrisierung wird in der Regel so gewählt, daß θ_0 im Innern von Θ liegt.

K6: Die Funktion $\pi(\cdot) = (\pi_1(\cdot), \dots, \pi_k(\cdot))'$ ist stetig.

K7: Die Umkehrabbildung $\pi^{-1}(\cdot) : \mathcal{Y} \longrightarrow \Theta$ existiert und ist stetig.

Äquivalent zu K7 ist, daß für alle $\theta \in \Theta$ und $\varepsilon > 0$ ein $\delta > 0$ existiert, so daß für alle $\vartheta \in \Theta$ mit $\|\theta - \vartheta\| > \varepsilon$ die Ungleichung $\|\pi(\theta) - \pi(\vartheta)\| > \delta$ folgt (strenge Identifizierbarkeit von θ durch $\pi(\cdot)$).

Satz 5.4

Gelten die Modellvoraussetzung I, die Annahmen A und B über die Asymptotik sowie die Regularitätsbedingungen K2 - K7, so folgen

$$\mathbb{P}_{n^*} \{ \mathcal{D}(\mathbb{P}_n, \pi(\hat{\theta}_n), N_n) = \inf_{\theta \in \Theta} \mathcal{D}(\mathbb{P}_n, \pi(\theta), N_n) \} \longrightarrow 1$$

und

$$\hat{\theta}_n \xrightarrow{\mathbb{P}_n} \theta_0.$$

Beweis:

Wegen K5 und K6 ist $\mathcal{P}$ lokalkompakt und Satz 5.3 daher anwendbar. Beide Behauptungen folgen nun sofort aus den entsprechenden Behauptungen dort: die erste aufgrund der Surjektivität von $\pi(\cdot): \Theta \rightarrow \mathcal{P}$ und die zweite aufgrund der Stetigkeit von $\pi^{-1}(\cdot)$, da die stochastische Konvergenz unter einer stetigen Abbildung erhalten bleibt. □

Auch dieser Satz läßt sich leicht auf den Fall verallgemeinern, daß π_0 nicht im Modellraum $\mathcal{P}$ liegt.

6. Asymptotische Normalität

Für den Beweis der asymptotischen Normalität von $\hat{\pi}_n$ bzw. $\hat{\theta}_n$ benötigen wir zusätzlich die Modellbedingung II und die folgende Verschärfung von $\pi_{on} \rightarrow \pi_o$:

Annahme C:

Es existiert ein $\mu = (\mu_1, \dots, \mu_k)' \in \mathbb{R}^k$, so daß

$$\sqrt{S_n} (\pi_{on} - \pi_o) \xrightarrow{\mathbb{P}_n} \mu$$

gilt.

π_{on} muß wiederum nicht in $\mathcal{S}$ liegen, soll aber mit bestimmter Geschwindigkeit gegen den Punkt $\pi_o \in \mathcal{S}$ konvergieren.

Wir zeigen zunächst ein allgemeines Lemma.

Lemma 6.1

Sei $\{Z_n\}$ eine Folge von reellwertigen Zufallsvariablen auf $(\Omega, \mathcal{A})$ mit $Z_n \xrightarrow{\mathbb{P}_n} \infty$. Sei weiter $g(\cdot)$ eine beschränkte und meßbare Funktion $\mathbb{R} \rightarrow \mathbb{R}$ mit $\lim_{u \rightarrow \infty} g(u) = 0$. Dann gilt

$$\mathbb{E}_n\{g(Z_n)\} \rightarrow 0.$$

Ist $\{\mathcal{T}_n\}$ eine Folge von σ -Algebren mit $\mathcal{T}_n \subset \mathcal{A}$, so gilt außerdem

$$\mathbb{E}_n\{g(Z_n) | \mathcal{T}_n\} \xrightarrow{\mathbb{P}_n} 0.$$

Beweis:

Sei $|g(\cdot)|$ beschränkt durch $K > 0$. Wegen $\lim_{u \rightarrow \infty} g(u) = 0$ gibt es zu jedem $\varepsilon > 0$ ein $a \in \mathbb{R}$ mit $\sup_{u \geq a} |g(u)| \leq \varepsilon$. Damit erhalten wir

$$\begin{aligned} \mathbb{E}_n\{|g(Z_n)|\} &= \int_{\{Z_n \leq a\}} |g(Z_n)| d\mathbb{P}_n + \int_{\{Z_n > a\}} |g(Z_n)| d\mathbb{P}_n \\ &\leq K \mathbb{P}_n\{Z_n \leq a\} + \varepsilon, \end{aligned}$$

woraus wegen $\mathbb{P}_n\{Z_n \leq a\} \rightarrow 0$ und der Tatsache, daß $\varepsilon > 0$ beliebig ist, die erste Behauptung folgt.

Analog gilt

$$\begin{aligned} (6.1) \quad \mathbb{E}_n\{|g(Z_n)| \mid \tau_n\} &= \mathbb{E}_n\{|g(Z_n)| \mathbb{I}_{(-\infty, a]}(Z_n) \mid \tau_n\} + \\ &\quad + \mathbb{E}_n\{|g(Z_n)| \mathbb{I}_{(a, \infty)}(Z_n) \mid \tau_n\} \\ &\leq K \mathbb{P}_n\{Z_n \leq a \mid \tau_n\} + \varepsilon \quad \mathbb{P}_n\text{-f.s.} \end{aligned}$$

Wegen $\mathbb{P}_n\{Z_n \leq a\} = \mathbb{E}_n[\mathbb{P}_n\{Z_n \leq a \mid \tau_n\}] \geq \delta \mathbb{P}_n[\mathbb{P}_n\{Z_n \leq a \mid \tau_n\} \geq \delta]$ für alle $\delta > 0$ (Markoff - Ungleichung) und $\mathbb{P}_n\{Z_n \leq a\} \rightarrow 0$ gilt außerdem

$$\mathbb{P}_n\{Z_n \leq a \mid \tau_n\} \xrightarrow{\mathbb{P}_n} 0.$$

Hieraus folgt mit (6.1) und der Tatsache, daß $\varepsilon > 0$ beliebig ist, die zweite Behauptung. □

Das nachfolgende Lemma ist zentral für die Herleitung der asymptotischen Normalität.

Lemma 6.2

Sind die Modellvoraussetzungen I und II sowie die Annahmen A, B und C über die Asymptotik erfüllt, so gilt

$$\sqrt{S_n} (P_n - \pi_0) \xrightarrow{\mathcal{L}_n} \mathcal{N}(\mu, \Sigma)$$

mit

$$(6.2) \quad \Sigma := \text{diag} \left\{ \frac{\pi_{0i}(1-\pi_{0i})}{\eta_i} \right\}.$$

Beweis :

Sei $V_n = (V_{n1}, \dots, V_{nk})'$, $n \in \mathbb{N}$, mit

$$V_{ni} := \sqrt{N_{ni}} \frac{P_{ni} - \pi_{oni}}{\sqrt{\pi_{oni}(1 - \pi_{oni})}}, \quad n \in \mathbb{N}, i=1, \dots, k.$$

Wir zeigen zunächst

$$(6.3) \quad V_n \xrightarrow{\mathcal{L}_n} \mathcal{N}(0, I).$$

Sei vorläufig $i \in \{1, \dots, k\}$ fest. Nach Modellvoraussetzung I existiert für alle $m \in \mathbb{N}_0$, $n \in \mathbb{N}$ ein bedingtes Wahrscheinlichkeitsmaß $P_n\{D_{ni} \in \cdot | N_{ni} = m\}$ von D_{ni} gegeben $N_{ni} = m$, definiert durch

$$P_n\{D_{ni} = d | N_{ni} = m\} = \binom{m}{d} \pi_{oni}^d (1 - \pi_{oni})^{m-d} \quad \forall 0 \leq d \leq m < \infty.$$

Sei weiter $v \in \mathbb{R}$ fest und sei $\varphi(\cdot)$ irgendeine Abbildung $\mathbb{N} \rightarrow \mathbb{N}$. Wir zeigen zunächst

$$(6.4) \quad |P_{\varphi(m)}\{V_{\varphi(m)i} \leq v | N_{\varphi(m)i} = m\} - \Phi(v)| \rightarrow 0 \quad (m \rightarrow \infty),$$

wobei $P_n\{V_{ni} \in \cdot | N_{ni} = m\}$ das durch $P_n\{D_{ni} \in \cdot | N_{ni} = m\}$ induzierte Wahrscheinlichkeitsmaß ist und $\Phi(\cdot)$ die Verteilungsfunktion der standardisierten Normalverteilung ist. Mit Hilfe des zentralen Grenzwertsatzes von Ljapunoff (Billingsley (68), S. 44) sieht man, daß (6.4) erfüllt ist, wenn für unabhängige, $\mathcal{B}(1, \pi_{oni})$ -verteilte Zufallsvariable R_{nj} , $j=1, \dots, m$, $n \in \mathbb{N}$, die Beziehung

$$(6.5) \quad \frac{1}{[m \text{Var}\{R_{\varphi(m)1}\}]^{3/2}} \sum_{j=1}^m \mathbb{E}\{|R_{\varphi(m)j}|\}^3 \rightarrow 0 \quad (m \rightarrow \infty)$$

gilt. (6.5) gilt aber, da wegen $\pi_{oni} \rightarrow \pi_{oi}$ und $0 < \pi_{oi} < 1$ die Folge $\{\pi_{oni}\}$ gleichmäßig von 0 und 1 weg beschränkt ist.

Als nächstes zeigen wir, daß aus (6.4)

$$(6.6) \quad g_i(m) := \sup_{n \in \mathbb{N}} |P_n\{V_{ni} \leq v | N_{ni} = m\} - \Phi(v)| \rightarrow 0 \quad (m \rightarrow \infty)$$

folgt. Angenommen es gäbe ein $\varepsilon > 0$, so daß für unendlich viele verschiedene $m \in \mathbb{N}$ ein $n(m) \in \mathbb{N}$ mit

$$|\mathbb{P}_{n(m)}\{V_{n(m)i} \leq v | N_{n(m)i} = m\} - \Phi(v)| > \varepsilon$$

existiere, so erhalten wir, wenn wir im anderen Fall $n(m) = 1$ setzen, mit $\varphi(m) := n(m)$, $m \in \mathbb{N}$, einen Widerspruch zu (6.4).

Mit Lemma 6.1 folgt aus (6.6) und $N_{ni} \xrightarrow{\mathbb{P}_n} \infty$ (Annahmen A und B):

$$\begin{aligned} (6.7) \quad |\mathbb{P}_n\{V_{ni} \leq v\} - \Phi(v)| &= \int |\mathbb{P}_n\{V_{ni} \leq v | N_{ni}\} - \Phi(v)| d\mathbb{P}_n \\ &\leq \int g_i(N_{ni}) d\mathbb{P}_n \\ &\longrightarrow 0. \end{aligned}$$

Außerdem gilt nach Definition der bedingten Wahrscheinlichkeit für alle $n \in \mathbb{N}$ und $i=2, \dots, k$

$$\begin{aligned} (6.8) \quad \mathbb{P}_n\{V_{ni} \leq v | V_{n1}, \dots, V_{n,i-1}\} \\ = \int \mathbb{P}_n\{V_{ni} \leq v | N_{ni}, V_{n1}, \dots, V_{n,i-1}\} d\mathbb{P}_n\{\cdot | V_{n1}, \dots, V_{n,i-1}\} \\ \mathbb{P}_n\text{-f.s.}, \end{aligned}$$

wobei $\mathbb{P}_n\{\cdot | V_{n1}, \dots, V_{n,i-1}\}$ eine bedingte Wahrscheinlichkeitsverteilung zu $\mathbb{P}_n$ gegeben $(V_{n1}, \dots, V_{n,i-1})$ ist. (Die Existenz ist gesichert, da $V_{n1}, \dots, V_{n,i-1}$ diskrete Zufallsvariablen sind.) Da wegen Modellvoraussetzung II

$$\mathbb{P}_n\{V_{ni} \leq v | N_{ni}, V_{n1}, \dots, V_{n,i-1}\} = \mathbb{P}_n\{V_{ni} \leq v | N_{ni}\} \quad \mathbb{P}_n\text{-f.s.}$$

gilt, folgt mit (6.6) und (6.8)

$$\begin{aligned}
 & |\mathbb{P}_n\{V_{ni} \leq v | V_{n1}, \dots, V_{n,i-1}\} - \phi(v)| \\
 & \leq \int |\mathbb{P}_n\{V_{ni} \leq v | N_{ni}\} - \phi(v)| d\mathbb{P}_n\{\cdot | V_{n1}, \dots, V_{n,i-1}\} \\
 & \leq \int g_i(N_{ni}) d\mathbb{P}_n\{\cdot | V_{n1}, \dots, V_{n,i-1}\} \quad \mathbb{P}_n\text{-f.s.}
 \end{aligned}$$

und mit Lemma 6.1 erhalten wir hieraus

$$(6.9) \quad |\mathbb{P}_n\{V_{ni} \leq v | V_{n1}, \dots, V_{n,i-1}\} - \phi(v)| \xrightarrow{\mathbb{P}_n} 0.$$

Aus (6.7) und (6.9) erhalten wir nun folgendermaßen (6.3):
Seien $v_1, \dots, v_k \in \mathbb{R}$ beliebig und

$$G_i := \{V_{n1} \leq v_1, \dots, V_{n,i-1} \leq v_{i-1}\}, \quad i=2, \dots, k.$$

Dann folgt aufgrund der Definition der bedingten Wahrscheinlichkeit, aus (6.9) und aus dem Satz von der majorisierten Konvergenz (Loève (77), S. 126)

$$\begin{aligned}
 & |\mathbb{P}_n(\{V_{ni} \leq v_i\} \cap G_i) - \phi(v_i) \mathbb{P}_n(G_i)| \\
 & \leq \int_{G_i} |\mathbb{P}_n\{V_{ni} \leq v_i | V_{n1}, \dots, V_{n,i-1}\} - \phi(v_i)| d\mathbb{P}_n \\
 & \longrightarrow 0, \quad i=2, \dots, k.
 \end{aligned}$$

Mit Hilfe vollständiger Induktion folgt hieraus und aus (6.7)

$$\mathbb{P}_n\{V_{n1} \leq v_1, \dots, V_{nk} \leq v_k\} \longrightarrow \prod_{i=1}^k \phi(v_i)$$

und damit (6.3).

Die Behauptung des Satzes erhalten wir nun aus (6.3) und aus

$$\begin{aligned}
 \sqrt{S_n} (P_n - \pi_0) &= \sqrt{S_n} (P_n - \pi_{on}) + \mu + o_{\mathbb{P}_n}(1) \\
 &= \text{diag} \left\{ \sqrt{\frac{S_n \pi_{oni} (1 - \pi_{oni})}{N_{ni}}} \right\} V_n + \mu + o_{\mathbb{P}_n}(1) \\
 &= \text{diag} \left\{ \sqrt{\frac{\pi_{oni} (1 - \pi_{oni})}{\eta_i}} \right\} V_n + \mu + o_{\mathbb{P}_n}(1).
 \end{aligned}$$

□

Zunächst betrachten wir den Fall, daß ρ durch die Parametrisierung $\pi(\cdot) : \Theta \rightarrow \mathcal{P}$, $\Theta \subset \mathbb{R}^r$, $0 < r \leq k$, gegeben ist. Wir zeigen die asymptotische Normalität von $\hat{\theta}_n$ unter den folgenden Annahmen:

N1: Θ ist offen.

N2: Die Funktionen $g_i(\cdot, \cdot) := D_2 h_i(\cdot, \cdot)$, $i=1, \dots, k$, existieren auf $[0, 1] \times (0, 1)$ und sind auf $(0, 1) \times (0, 1)$ stetig differenzierbar. Außerdem gilt für alle $a \in (0, 1)$, $i=1, \dots, k$,

$$g_i(a, a) = 0,$$

$$D_2 g_i(a, a) > 0,$$

$$D_1 g_i(a, a) \neq 0.$$

Die Bedingungen $g_i(a, a) = 0$ und $D_2 g_i(a, a) > 0$ folgen aus K3, sofern $h_i(\cdot)$ die nötigen Differenzierbarkeitseigenschaften besitzt. Die Bedingung $D_1 g_i(a, a) \neq 0$ wird benötigt, damit die Limesverteilung von $\hat{\theta}_n$ nicht auf eine Hyperebene im $\mathbb{R}^r$ beschränkt ist.

N3: Die Funktion $\pi(\cdot) = (\pi_1(\cdot), \dots, \pi_k(\cdot))'$ ist stetig differenzierbar.

N4: Die $k \times r$ -Matrix $D\pi(\theta) = \left(\frac{\partial \pi_i}{\partial \theta_j}(\theta) \right)$ besitzt für alle $\theta \in \Theta$ vollen Rang r .

Aus der Gültigkeit von N1, N3 und N4 folgt, daß $\{\pi(\theta) : \theta \in \Theta\}$ lokal eine C^1 -Mannigfaltigkeit im $\mathbb{R}^r$ ist (Satz A1 im Anhang A). Gilt zusätzlich K7, so ist $\{\pi(\theta) : \theta \in \Theta\}$ global eine C^1 -Mannigfaltigkeit im $\mathbb{R}^r$ (Satz A2 im Anhang A).

Im folgenden benutzen wir die $r \times r$ - und $k \times r$ -Matrizen

$$(6.10) \quad J(\theta, q) := D\pi(\theta)' \text{diag}\{q_i D_2 g_i(\pi_i(\theta), \pi_i(\theta))\} D\pi(\theta)$$

und

$$(6.11) \quad A(\theta, q) := \text{diag}\{q_i D_1 g_i(\pi_i(\theta), \pi_i(\theta))\}' D\pi(\theta)$$

sowie den r -dimensionalen Zufallsvektor

$$Z_n(\theta) = (Z_{n1}(\theta), \dots, Z_{nr}(\theta))',$$

definiert durch

$$(6.12) \quad Z_{nj}(\theta) := \frac{1}{S_n} \frac{\partial}{\partial \theta_j} \mathcal{D}(P_n, \pi(\theta), N_n) \\ = \frac{1}{S_n} \sum_{i=1}^k N_{ni} g_i(P_{ni}, \pi_i(\theta)) \frac{\partial \pi_i}{\partial \theta_j}(\theta), \quad j=1, \dots, r.$$

Außerdem benutzen wir die Abkürzungen $J := J(\theta_0, \eta)$, $A := A(\theta_0, \eta)$ und $Z_n := Z_n(\theta_0)$. Aus den Annahmen B und N2 - N4 folgt insbesondere, daß J und $A'A$ positiv definit und daher invertierbar sind.

Für den nachfolgenden Satz muß $\hat{\theta}_n$ nicht unbedingt ein Schätzer bzgl. $\mathcal{D}(\cdot)$ (wie im Abschnitt 4 definiert) sein, sondern es reicht, wenn $\hat{\theta}_n$ irgendeine (nicht notwendig meßbare) Variable $\Omega \rightarrow \Theta$ mit den Eigenschaften

$$(6.13) \quad \hat{\theta}_n \xrightarrow{\mathbb{P}_n} \theta_0$$

und

$$(6.14) \quad \frac{1}{\sqrt{S_n}} \frac{\partial}{\partial \theta_j} \mathcal{D}(P_n, \pi(\hat{\theta}_n), N_n) \xrightarrow{\mathbb{P}_n} 0, \quad j=1, \dots, r,$$

ist. Gelten N1, N2, N3 und die Voraussetzungen des Satzes 5.4, so erfüllt jeder Schätzer $\hat{\theta}_n$ bzgl. $\mathcal{D}(\cdot)$ die Bedingung (6.14), da wegen

$$\{\mathcal{D}(P_n, \pi(\hat{\theta}_n), N_n) = \inf_{\theta \in \Theta} \mathcal{D}(P_n, \pi(\theta), N_n)\} \\ \subset \left\{ \frac{\partial}{\partial \theta_j} \mathcal{D}(P_n, \pi(\hat{\theta}_n), N_n) = 0, \quad j=1, \dots, r \right\}$$

die Beziehungen

$$(6.15) \quad \mathbb{P}_n * \left\{ \frac{\partial}{\partial \theta_j} \mathcal{D}(P_n, \pi(\hat{\theta}_n), N_n) = 0 \right\} \longrightarrow 1, \quad j=1, \dots, r,$$

gelten.

Satz 6.3

Gelten die Modellannahmen I und II, die Annahmen A, B und C sowie die Regularitätsvoraussetzungen N1 - N3, so folgen

$$(6.16) \quad \sqrt{S_n} Z_n - A' \sqrt{S_n} (P_n - \pi_0) \xrightarrow{\mathbb{P}_n} 0$$

und

$$(6.17) \quad \sqrt{S_n} Z_n \xrightarrow{\mathcal{L}_n} \mathcal{N}(A' \mu, A' \Sigma A),$$

wobei Σ durch (6.2) definiert ist. Erfüllt zusätzlich die Folge $\{\hat{\theta}_n\}$ die Bedingungen (6.13) und (6.14) und gilt N4, so folgen außerdem

$$(6.18) \quad \sqrt{S_n} (\hat{\theta}_n - \theta_0) + J^{-1} \sqrt{S_n} Z_n \xrightarrow{\mathbb{P}_n} 0$$

und

$$(6.19) \quad \sqrt{S_n} (\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}_n} \mathcal{N}(-J^{-1} A' \mu, J^{-1} A' \Sigma A J^{-1}).$$

Beweis :

Für jedes $i \in \{1, \dots, k\}$ läßt sich $g_i(\cdot)$ wegen N2 nach Taylor entwickeln, so daß mit $P_{ni} \xrightarrow{\mathbb{P}_n} \pi_{oi} = \pi_i(\theta_0)$ (Lemma 5.1) und $g_i(\pi_{oi}, \pi_{oi}) = 0$ (N2)

$$g_i(P_{ni}, \pi_{oi}) = D_1 g_i(\pi_{oi}, \pi_{oi}) (P_{ni} - \pi_{oi}) + |P_{ni} - \pi_{oi}| o_{\mathbb{P}_n}(1)$$

folgt. Nach Multiplikation beider Seiten mit $\frac{N_{ni}}{\sqrt{S_n}} \frac{\partial \pi_i}{\partial \theta_j}(\theta_0)$, $j=1, \dots, r$, und Summation über $i=1, \dots, k$ erhalten wir hieraus in Matrixschreibweise

$$\begin{aligned} \sqrt{S_n} Z_n &= D\pi(\theta_0)' \text{diag}\left\{\frac{N_{ni}}{S_n} D_1 g_i(\pi_{oi}, \pi_{oi})\right\} \sqrt{S_n} (P_n - \pi_0) + \\ &+ \left\| \text{diag}\left\{\frac{N_{ni}}{S_n}\right\} \sqrt{S_n} (P_n - \pi_0) \right\| o_{\mathbb{P}_n}(1). \end{aligned}$$

Da $N_n/S_n \xrightarrow{\mathbb{P}_n} \eta$ (Annahme B) gilt und da $\sqrt{S_n} (P_n - \pi_0)$ eine Limesverteilung besitzt (Lemma 6.2), erhalten wir somit

$$\sqrt{S_n} Z_n = A' \sqrt{S_n} (P_n - \pi_0) + o_{P_n}(1).$$

Damit ist (6.16) bewiesen. (6.17) ist eine Folgerung aus (6.16) und Lemma 6.2.

Weiter liefert für jedes $i \in \{1, \dots, k\}$ die Taylorentwicklung von $g_i(\cdot, \pi_i(\cdot))$ um $(\pi_{0i}, \pi_i(\theta_0))$ zusammen mit $g_i(\pi_{0i}, \pi_{0i}) = 0$, $P_{ni} \xrightarrow{P_n} \pi_{0i}$ und $\hat{\theta}_n \xrightarrow{P_n} \theta_0$

$$\begin{aligned} g_i(P_{ni}, \hat{\pi}_{ni}) &= [D_1 g_i(\pi_{0i}, \pi_{0i}) + o_{P_n}(1)] (P_{ni} - \pi_{0i}) + \\ &+ \sum_{s=1}^r [D_2 g_i(\pi_{0i}, \pi_{0i}) \frac{\partial \pi_i}{\partial \theta_s}(\theta_0) + o_{P_n}(1)] (\hat{\theta}_{ns} - \theta_0) \end{aligned}$$

Nach Multiplikation beider Seiten mit $\frac{N_{ni}}{\sqrt{S_n}} \frac{\partial \pi_i}{\partial \theta_t}(\hat{\theta}_n)$, $t=1, \dots, r$, folgt hieraus unter Berücksichtigung von $D\pi(\hat{\theta}_n) \xrightarrow{P_n} D\pi(\theta_0)$ und $N_n/S_n \xrightarrow{P_n} \eta$

$$\begin{aligned} \sqrt{S_n} Z_n(\hat{\theta}_n) &= [A' + o_{P_n}(1)] \sqrt{S_n} (P_n - \pi_0) + \\ &+ [J + o_{P_n}(1)] \sqrt{S_n} (\hat{\theta}_n - \theta_0) \end{aligned}$$

und weiter mit (6.14) und Lemma 6.2

$$A' \sqrt{S_n} (P_n - \pi_0) = - [J + o_{P_n}(1)] \sqrt{S_n} (\hat{\theta}_n - \theta_0) + o_{P_n}(1).$$

Da eine Determinante als stetige Funktion ihrer Elemente aufgefaßt werden kann und J invertierbar ist, folgt somit

$$\begin{aligned} \sqrt{S_n} (\hat{\theta}_n - \theta_0) &= - [J^{-1} + o_{P_n}(1)] A' \sqrt{S_n} (P_n - \pi_0) + o_{P_n}(1) \\ &= - J^{-1} A' \sqrt{S_n} (P_n - \pi_0) + o_{P_n}(1). \end{aligned}$$

Mit (6.16) erhalten wir daher (6.18). (6.19) ergibt sich sofort aus (6.17) und (6.18).



Im Fall, daß die Abstandsfunktionen $h_i(\cdot)$, $i=1, \dots, k$, durch (4.6), (4.8) oder (4.9) definiert sind, gilt

$$(6.20) \quad D_2 g_i(a, a) = -D_1 g_i(a, a) = \frac{c}{a(1-a)} \quad \forall a \in (0, 1), \quad i=1, \dots, k,$$

wobei beim Maximum-Likelihood-Schätzer $c=1$ und bei den beiden Minimum- χ^2 -Schätzern $c=2$ ist. Die Matrizen J und A sind dann von der Form $J=cJ^*$ und $A=cA^*$ mit

$$J^* := D\pi(\theta_0)' \Sigma^{-1} D\pi(\theta_0)$$

und

$$A^* := -\Sigma^{-1} D\pi(\theta_0).$$

Außerdem liefern (6.16) und (6.18)

$$\sqrt{S_n}(\hat{\theta}_n - \theta_0) + J^{*-1} A^{*'} \sqrt{S_n}(P_n - \pi_0) \xrightarrow{\mathbb{P}_n} 0.$$

Man sieht hieraus, daß alle drei Schätzer in dem Sinne asymptotisch äquivalent sind, daß für je zwei, etwa $\hat{\theta}_n$ und $\tilde{\theta}_n$, die Beziehung

$$\sqrt{S_n}(\hat{\theta}_n - \tilde{\theta}_n) \xrightarrow{\mathbb{P}_n} 0$$

gilt. Insbesondere besitzen die Schätzer damit die gleiche Limesverteilung:

$$\sqrt{S_n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}_n} \mathcal{N}(-J^{*-1} A^{*'} \mu, J^{*-1}),$$

wie sich sofort aus (6.19) ergibt. Falls die partielle log-Likelihoodfunktion (4.5) bis auf eine Konstante gleich der vollen log-Likelihoodfunktion ist, läßt sich folgendes sagen:

- jeder der drei Schätzer ist effizient erster Ordnung (Rao (73), S. 348),
- bzgl. der Effizienz zweiter Ordnung ist der Maximum-Likelihood-Schätzer dem Minimum- χ^2 -Schätzer und dieser dem modifizierten Minimum- χ^2 -Schätzer vorzuziehen (Rao (63)).

Wir haben oben gesehen, daß jeder konsistente Schätzer $\hat{\theta}_n$ bzgl. $\mathcal{D}(\cdot)$ unter den Voraussetzungen des Satzes 5.4 und unter N1-N3 die Bedingungen (6.13) und (6.15) erfüllt. Wir geben nun Bedingungen an, unter denen für jede Folge $\{\hat{\theta}_n\}$, die (6.13) und (6.15) erfüllt, $\hat{\theta}_n$ nach Wahrscheinlichkeit eindeutig ist. Insbesondere ist damit gezeigt, daß für jede konsistente Folge $\{\hat{\theta}_n\}$, die (6.15) erfüllt, $\hat{\theta}_n$ nach Wahrscheinlichkeit gleich dem Schätzer bzgl. $\mathcal{D}(\cdot)$ ist. Für diesen Satz werden wie beim Konsistenzbeweis die Modellannahme II und die Bedingung C nicht benötigt.

Satz 6.4

Gelten die Modellannahme I, die Annahmen A und B über die Asymptotik sowie die Regularitätsannahmen N1-N4. Sei zusätzlich $\pi(\cdot)$ zweimal stetig differenzierbar. Dann ist für jede Folge $\{\hat{\theta}_n\}$ von Variablen, für die

$$\hat{\theta}_n \xrightarrow{\mathbb{P}_n} \theta_0$$

und

$$\mathbb{P}_{n*} \left\{ \frac{\partial}{\partial \theta_j} \mathcal{D}(\mathbb{P}_n, \pi(\hat{\theta}_n), N_n) = 0 \right\} \longrightarrow 1, \quad j=1, \dots, r,$$

gelten, $\hat{\theta}_n$ nach $\mathbb{P}_n$ -Wahrscheinlichkeit eindeutig. +)

Beweis:

Sei $f(p, q, \theta) := \mathcal{D}(p, \pi(\theta), q)$. Aus den Voraussetzungen folgt, daß $D_3 f(\cdot, \cdot, \cdot)$ und $D_3^2 f(\cdot, \cdot, \cdot)$ existieren und stetig sind. Außerdem gilt wegen N2 und N3 $D_3 f(\pi_0, \eta, \theta_0) = 0$ und wegen N2-N4 und $\eta > 0$ besitzt die Matrix $D_3^2 f(\pi_0, \eta, \theta_0)$ vollen Rang r . Damit sind die Voraussetzungen für den Satz über implizite Funktionen (Loomis, Sternberg (68), S.230) erfüllt

+) Sei $\{A_n\}$ eine Folge nicht notwendig meßbarer Ereignisse ($A_n \subset \Omega$). Dann gelte die Definition: A_n gilt nach $\mathbb{P}_n$ -Wahrscheinlichkeit : $\iff \mathbb{P}_{n*}(A_n) \rightarrow 1$. ($\mathbb{P}_{n*}$ ist das zu $\mathbb{P}_n$ gehörende innere Wahrscheinlichkeitsmaß; siehe Anhang B.)

und es gibt offene Umgebungen U um (π_0, η) und $V \subset \Theta$ um θ_0 und eine eindeutige Funktion $u(\cdot): U \rightarrow V$ mit

$$D_3 f(p, q, u(p, q)) = 0 \quad \forall (p, q) \in U.$$

Definiere $u(\cdot): [0, 1]^{2k} \rightarrow \Theta$ beliebig außerhalb U . Da wegen Annahme B und Lemma 5.1 $(P_n, N_n/S_n)$ nach $\mathbb{P}_n$ -Wahrscheinlichkeit in U liegt, gilt daher

$$\mathbb{P}_n * \{D_3 f(P_n, \frac{N_n}{S_n}, u(P_n, \frac{N_n}{S_n})) = 0\} \longrightarrow 1.$$

Da außerdem nach Voraussetzung

$$\mathbb{P}_n * \{D_3 f(P_n, \frac{N_n}{S_n}, \hat{\theta}_n) = 0\} \longrightarrow 1$$

gilt und $u(P_n, N_n/S_n)$ nach $\mathbb{P}_n$ -Wahrscheinlichkeit eindeutig ist, folgt somit

$$\mathbb{P}_n * \{\hat{\theta}_n = u(P_n, \frac{N_n}{S_n})\} \longrightarrow 1$$

und damit die Behauptung. □

Wir kommen nun zu dem Fall, bei dem $\mathcal{J}$ nicht explizit durch eine Parametrisierung gegeben ist. Mit Hilfe des Satzes 6.3 und des Konzepts der C^1 -Mannigfaltigkeiten (siehe Anhang A) lassen sich die Bedingungen für die asymptotische Normalität von folgendermaßen formulieren:

Satz 6.5

Gelten die Modellannahmen I und II und die Annahmen A, B und C über die Asymptotik. Sei $\mathcal{Y}$ eine r -dimensionale C^1 -Mannigfaltigkeit in $(0,1)^k$ und erfülle $\mathcal{D}(\cdot)$ die Regularitätsbedingungen N2, K3 und K4. Ist $\{\hat{\pi}_n\}$ eine Folge von Schätzern bzgl. $\mathcal{D}(\cdot)$, so gelten

$$(6.21) \quad \sqrt{S_n}(\hat{\pi}_n - \pi_0) + G J^{-1} \bar{A}' \sqrt{S_n}(P_n - \pi_0) \xrightarrow{\mathbb{P}_n} 0$$

und

$$(6.22) \quad \sqrt{S_n}(\hat{\pi}_n - \pi_0) \xrightarrow{\mathcal{L}_n} \mathcal{N}(-G J^{-1} \bar{A}' \mu, G J^{-1} \bar{A}' \Sigma \bar{A} J^{-1} G'),$$

wobei G irgendeine $k \times r$ -Matrix ist, deren Spalten den Tangentialraum in π_0 an $\mathcal{Y}$ aufspannen, und die Matrizen $\bar{A}$ und J durch

$$\bar{A} := \text{diag}\{\eta_i D_1 g_i(\pi_{oi}, \pi_{oi})\} G,$$

$$J := G' \text{diag}\{\eta_i D_2 g_i(\pi_{oi}, \pi_{oi})\} G$$

definiert sind.

Beweis:

Nach Satz A3 im Anhang A gibt es eine offene Umgebung U um π_0 und eine invertierbare und stetig differenzierbare Funktion $\pi(\cdot)$ von einer offenen Teilmenge $\Theta \subset \mathbb{R}^r$ nach $U \cap \mathcal{Y}$, so daß die Matrix $D\pi(\theta)$ für alle $\theta \in \Theta$ vollen Rang r besitzt und die Umkehrabbildung $\pi^{-1}(\cdot): U \cap \mathcal{Y} \rightarrow \Theta$ stetig ist. Sei $\theta_0 := \pi^{-1}(\pi_0)$. Nach Satz 5.4 folgt dann $\hat{\theta}_n \xrightarrow{\mathbb{P}_n} \theta_0$. Außerdem gilt (6.14), da Θ offen ist, $\pi(\cdot)$ stetig ist und N2 erfüllt ist. Damit sind alle Voraussetzungen für den Satz 6.3 erfüllt und wir erhalten aus (6.16) und (6.18)

$$(6.23) \quad \sqrt{S_n}(\hat{\theta}_n - \theta_0) + J^{-1} \bar{A}' \sqrt{S_n}(P_n - \pi_0) \xrightarrow{\mathbb{P}_n} 0,$$

wobei $G := D\pi(\theta_0)$ gesetzt ist.

Weiter liefert eine Taylorentwicklung

$$(6.24) \quad \sqrt{S_n} (\pi(\hat{\theta}_n) - \pi_0) - \sqrt{S_n} G (\hat{\theta}_n - \theta_0) \xrightarrow{\mathbb{P}_n} 0.$$

Da nach Satz 5.3

$$P_n * \left\{ \inf_{\pi \in \mathcal{J} \cap U} \mathcal{D}(P_n, \pi, N_n) = \inf_{\pi \in \mathcal{J}} \mathcal{D}(P_n, \pi, N_n) \right\} \longrightarrow 1$$

und damit

$$P_n * \{ \hat{\pi}_n = \pi(\hat{\theta}_n) \} \longrightarrow 1$$

gelten, folgt aus (6.24)

$$\sqrt{S_n} (\hat{\pi}_n - \pi_0) - \sqrt{S_n} G (\hat{\theta}_n - \theta_0) \xrightarrow{\mathbb{P}_n} 0.$$

Mit Lemma 6.2 ergeben sich hieraus und aus (6.23) die Gleichungen (6.21) und (6.22) für dieses spezielle G . Um zu zeigen, daß die Behauptung auch für irgendeine den Tangentialraum aufspannende $k \times r$ -Matrix gilt, genügt es zu beweisen, daß (6.21) und (6.22) unverändert bleiben, wenn G von rechts mit einer invertierbaren $r \times r$ -Matrix multipliziert wird. Dies ist leicht nachzurechnen. □

Gelten - wie bei den in Abschnitt 4 vorgestellten Abstandsfunktionen -

$$D_2 g_i(\pi_{oi}, \pi_{oi}) = -D_1 g_i(\pi_{oi}, \pi_{oi}), \quad i=1, \dots, k,$$

so läßt sich (6.21) schreiben als

$$\sqrt{S_n} (\hat{\pi}_n - \pi_0) = P_G \sqrt{S_n} (P_n - \pi_0) + o_{\mathbb{P}_n}(1)$$

oder

$$\sqrt{S_n} P_G (P_n - \hat{\pi}_n) \xrightarrow{\mathbb{P}_n} 0,$$

wobei $P_G := G(G'HG)^{-1}G'H$ mit $H := \text{diag}\{\eta_i D_2 g_i(\pi_{oi}, \pi_{oi})\}$ die orthogonale Projektion auf den von den Spalten von G aufgespannten Raum $\mathcal{M}(G)$ bzgl. des Skalarprodukts definiert durch $\langle X, Y \rangle := Y'HX$ ist.

7. Testen von Hypothesen

Seien $\mathcal{Y}$ und $\mathcal{Y}'$ eine r - und eine s -dimensionale C^1 -Mannigfaltigkeit in $(0,1)^k$ mit $0 \leq s < r \leq k$ und $\mathcal{Y}' \subset \mathcal{Y}$. Getestet werden soll die Nullhypothese

$$H_0: \pi \in \mathcal{Y}'$$

gegen die Alternative

$$H_1: \pi \in \mathcal{Y} - \mathcal{Y}'.$$

Im Fall, daß $\mathcal{Y}'$ einelementig ist, heißt die Nullhypothese einfach, im Fall $s > 0$ zusammengesetzt. Für $\mathcal{Y} = (0,1)^k$ erhält man einen globalen Anpassungstest für $\mathcal{Y}'$.

Bei Anwendungen ist $\mathcal{Y}$ meist von der Form

- (a) $\mathcal{Y} := \{\pi(\theta) : \theta \in \Theta\}$, wobei $\Theta \subset \mathbb{R}^r$ offen ist und $\pi(\cdot)$ eine injektive, stetig differenzierbare Funktion $\Theta \rightarrow \mathbb{R}^k$ ist, deren Umkehrabbildung $\pi^{-1}(\cdot) : \pi(\Theta) \rightarrow \Theta$ stetig ist und deren Ableitung $D\pi(\theta)$ für alle $\theta \in \Theta$ vollen Rang r besitzt (siehe Satz A2).

$\mathcal{Y}'$ ist dann meist auf eine der beiden folgenden Arten gegeben:

- (b) $\mathcal{Y}' := \{\pi(\theta) : \theta \in \Theta, R(\theta) = 0\}$, wobei $R(\cdot)$ eine stetig differenzierbare Funktion $\Theta \rightarrow \mathbb{R}^{r-s}$ ist und die $(r-s) \times r$ -Matrix $DR(\theta)$ für alle $\theta \in \Theta$ den Rang $r-s$ besitzt (siehe Sätze A4 und A7).
- (c) $\mathcal{Y}' := \{\pi \circ f(\beta) : \beta \in \Theta'\}$, wobei $\Theta' \subset \mathbb{R}^s$ offen ist und $f(\cdot)$ eine injektive, stetig differenzierbare Funktion $\Theta' \rightarrow \Theta$ ist, deren Umkehrabbildung $f^{-1}(\cdot) : f(\Theta') \rightarrow \Theta'$ stetig ist und deren Ableitung $Df(\beta)^{+}$ für alle $\beta \in \Theta'$ den Rang s besitzt (siehe Sätze A2 und A7).

Umgekehrt gibt es bei vorgegebenen r - und s -dimensionalen C^1 -Mannigfaltigkeiten $\mathcal{Y}$ und $\mathcal{Y}'$ mit $\mathcal{Y}' \subset \mathcal{Y}$, $0 \leq s < r \leq k$, zu jedem Punkt aus $\mathcal{Y}'$ eine Umgebung U und Funktionen $\pi(\cdot)$, $R(\cdot)$ und $f(\cdot)$,

+) $Df(\beta)$ sei im Fall $s = 0$ als $r \times 1$ -Matrix definiert, die nur Nullen enthält.

so daß (a), (b) und (c) erfüllt sind, wenn dort $\mathcal{Y}$ und $\mathcal{Y}'$ durch $\mathcal{Y} \cap U$ und $\mathcal{Y}' \cap U$ ersetzt werden. (Dies folgt aus den Sätzen A3, A5 und A6.) $\mathcal{Y}$ und $\mathcal{Y}'$ besitzen also lokal die Charakterisierungen (a), (b) und (c).

Sei im folgenden wieder $\mathcal{D}(\cdot)$ eine Funktion, die (4.1) erfüllt. Wir diskutieren in diesem Abschnitt drei Tests:

Test I:

Dieser Test wird benutzt, wenn $\mathcal{Y}$ und $\mathcal{Y}'$ durch (a) und (b) gegeben sind. Sei $\hat{\theta}_n$ der Schätzer bzgl. $\mathcal{D}(\cdot)$ für $\theta \in \Theta = \pi^{-1}(\mathcal{Y})$, seien $\hat{A}_n := A(\hat{\theta}_n, N_n/S_n)$ durch (6.11) und $\hat{J}_n := J(\hat{\theta}_n, N_n/S_n)$ durch (6.10) definiert und gelten $\hat{R}_n := R(\hat{\theta}_n)$, $\hat{T}_n := DR(\hat{\theta}_n)$ und $\hat{\Sigma}_n := \text{diag}\{\hat{\pi}_{ni}(1-\hat{\pi}_{ni})S_n/N_{ni}\}$ mit $\hat{\pi}_n := \pi(\hat{\theta}_n)$. Dann ist die Teststatistik W_n derart definiert, daß

$$(7.1) \quad W_n := S_n \hat{R}_n' (\hat{T}_n \hat{J}_n^{-1} \hat{A}_n' \hat{\Sigma}_n \hat{A}_n \hat{J}_n^{-1} \hat{T}_n)^{-1} \hat{R}_n$$

gilt, sofern die rechte Seite definiert ist. Für diesen Test wird nur der Schätzer unter dem Modell $\mathcal{Y}$ benötigt. Die Teststatistik ist von der Wahl der Funktionen $\pi(\cdot)$ und $R(\cdot)$ abhängig.

Test II:

Dieser Test wird benutzt, wenn $\mathcal{Y}$ und $\mathcal{Y}'$ durch (a) und (c) gegeben sind. Sei $\hat{\theta}_n$ der Schätzer bzgl. $\mathcal{D}(\cdot)$ für $\theta \in \pi^{-1}(\mathcal{Y}')$, seien $\hat{A}_n := A(\hat{\theta}_n, N_n/S_n)$ durch (6.11) und $\hat{Z}_n := Z_n(\hat{\theta}_n)$ durch (6.12) definiert und gelte $\hat{\Sigma}_n := \text{diag}\{\hat{\pi}_{ni}(1-\hat{\pi}_{ni})S_n/N_{ni}\}$ mit $\hat{\pi}_n := \pi(\hat{\theta}_n)$. Dann ist die Teststatistik C_n derart definiert, daß

$$(7.2) \quad C_n := S_n \hat{Z}_n' (\hat{A}_n' \hat{\Sigma}_n \hat{A}_n)^{-1} \hat{Z}_n$$

gilt, sofern die rechte Seite definiert ist. Für diesen Test wird nur der Schätzer unter der Nullhypothese benötigt. Die Teststatistik hängt nicht von der Wahl der Parametrisierung $\pi(\cdot)$ ab, da (7.2) äquivalent geschrieben werden kann als

$$C_n := S_n \hat{M}_n' \hat{G}_n (\hat{G}_n' \hat{H}_n \hat{G}_n)^{-1} \hat{G}_n' \hat{M}_n$$

wobei $\hat{H}_n := \text{diag}\{[D_1 g_1(\hat{\pi}_{n1}, \hat{\pi}_{n1})]^2 \hat{\pi}_{n1} (1 - \hat{\pi}_{n1}) N_{n1} / S_n\}$,
 $\hat{M}_n := (g_1(P_{n1}, \hat{\pi}_{n1}) N_{n1} / S_n, \dots, g_k(P_{nk}, \hat{\pi}_{nk}) N_{nk} / S_n)'$ gelten und
 $\hat{G}_n$ irgendeine $k \times r$ -Matrix ist, deren Spalten den Tangentialraum in $\hat{\pi}_n$ an $\mathcal{Y}$ aufspannen.

Test III:

Die Teststatistik lautet

$$(7.3) \quad \Lambda_n := \mathcal{D}(P_n, \hat{\pi}_n, N_n) - \mathcal{D}(P_n, \hat{\pi}_n, N_n),$$

wobei $\hat{\pi}_n$ und $\hat{\pi}_n$ die Schätzer bzgl. $\mathcal{D}(\cdot)$ für $\pi \in \mathcal{Y}$ und $\pi \in \mathcal{Y}'$ sind.

Falls $\mathcal{D}(\cdot)$ bis auf eine Konstante die volle log-Likelihoodfunktion ist, sind W_n , C_n und Λ_n gerade die Teststatistiken für den Wald-Test, den Score-Test und den Likelihood-Quotienten-Test.

Wir leiten die asymptotische Verteilung dieser Teststatistiken wieder für den nichtzentralen Fall her, also unter der Annahme C, daß der wahre Parameter π_{on} mit bestimmter Geschwindigkeit gegen einen Punkt π_o konvergiert. Hierbei müssen die π_{on} nicht notwendig in $\mathcal{Y}'$ oder $\mathcal{Y}$ liegen, jedoch soll $\pi_o \in \mathcal{Y}'$ gelten. Die asymptotische Verteilung von W_n berechnen wir für den Fall, daß $\hat{\pi}_n$ bzw. $\hat{\theta}_n$ der Schätzer bzgl. einer Funktion $\mathcal{D}(\cdot)$ ist, die den Regularitätsbedingungen des vorhergehenden Abschnitts genügt. Die asymptotische Verteilung von C_n und Λ_n bestimmen wir nur für Abstandsfunktionen, für die zusätzlich ein $c > 0$ mit $A(\theta_o, \eta) \sum A(\theta_o, \eta) = c J(\theta_o, \eta)$ existiert, wobei $\sum$ durch (6.2) definiert ist. Wir schreiben wieder abkürzend P_n für das Wahrscheinlichkeitsmaß $P_{\pi_{on}}$ und entsprechend E_n , Var_n und $\mathcal{L}_n$ für $E_{\pi_{on}}$, $Var_{\pi_{on}}$ und $\mathcal{L}_{\pi_{on}}$.

Bei den folgenden beiden Sätzen nehmen wir o.B.d.A. an, daß $\mathcal{Y}$ und $\mathcal{Y}'$ die Charakterisierungen (a), (b) und (c) besitzen. Nach der obigen Bemerkung gibt es nämlich eine Umgebung U um π_o , so daß $\mathcal{Y} \cap U$ und $\mathcal{Y}' \cap U$ (a), (b) und (c) erfüllen. Da die Schätzer für $\pi \in \mathcal{Y} \cap U$ und $\pi \in \mathcal{Y}$ bzw. für $\pi \in \mathcal{Y}' \cap U$ und $\pi \in \mathcal{Y}'$ unter den Voraussetzungen für Satz 5.3 nach Wahrscheinlichkeit gleich sind, ändern sich ihre asymptotischen Eigenschaften nicht, wenn in den Sätzen $\mathcal{Y}$ und $\mathcal{Y}'$ durch $\mathcal{Y} \cap U$ und $\mathcal{Y}' \cap U$ ersetzt werden.

Satz 7.1

Seien $\mathcal{P}$ und $\mathcal{P}'$ eine r - und eine s -dimensionale C^1 -Mannigfaltigkeit in $(0,1)^k$ mit $0 \leq s < r \leq k$ und $\mathcal{P}' \subset \mathcal{P}$, welche die Charakterisierungen (a), (b) und (c) besitzen. Gelten die Modellannahmen I und II und die Annahmen A, B und C mit $\pi_0 \in \mathcal{P}'$ und erfülle $h_i(\cdot)$, $i=1, \dots, k$, die Regularitätsbedingungen K2, K3, K4 und N2. Dann folgt

$$W_n \xrightarrow{\mathcal{L}_n} \chi^2_{r-s}(\mu' A J^{-1} T' (T J^{-1} A' \Sigma A J^{-1} T')^{-1} T J^{-1} A' \mu),$$

wobei $J := J(\theta_0, \eta)$, $A := A(\theta_0, \eta)$ und $T := DR(\theta_0)$ mit $\theta_0 := \pi^{-1}(\pi_0)$ gelten und $\chi^2_t(\delta)$ die nichtzentrale χ^2 -Verteilung mit t Freiheitsgraden und Nichtzentralitätsparameter δ ist.

Beweis:

Wegen (a) sind die Regularitätsvoraussetzungen K5-K7 erfüllt. Außerdem gelten nach Voraussetzung K2-K4. Nach Satz 5.4 erhalten wir daher

$$\hat{\theta}_n \xrightarrow{P_n} \theta_0.$$

Da zusätzlich wegen (a) N1, N3 und N4 und nach Voraussetzung N2 erfüllt sind, folgt somit mit Satz 6.3

$$\sqrt{S}_n (\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}_n} \mathcal{N}(-J^{-1} A' \mu, J^{-1} A' \Sigma A J^{-1}).$$

Mit Berücksichtigung von $R(\theta_0) = 0$ erhalten wir hieraus

$$(7.4) \quad \sqrt{S}_n \hat{R}_n \xrightarrow{\mathcal{L}_n} \mathcal{N}(-T J^{-1} A' \mu, T J^{-1} A' \Sigma A J^{-1} T').$$

Definiere weiter $M := A J^{-1} T'$ und $\hat{M}_n := \hat{A}_n \hat{J}_n^{-1} \hat{T}'_n$. Da Σ , A , J und T vollen Rang besitzen, ist $M' \Sigma M$ invertierbar. Damit folgt aus

$$\mathcal{L}\{X\} = \mathcal{N}(-M' \mu, M' \Sigma M)$$

die Beziehung

$$(7.5) \quad \mathcal{L}\{X'(M'\Sigma M)^{-1}X\} = \chi^2_{r-s}(\mu'M(M'\Sigma M)^{-1}M'\mu)$$

(Rao(73), S.188). Da $X'(M'\Sigma M)^{-1}X$ eine stetige Funktion von X ist, erhalten wir aus (7.4) und (7.5)

$$(7.6) \quad S_n \hat{R}'_n (M'\Sigma M)^{-1} \hat{R}_n \xrightarrow{\mathbb{P}_n} \chi^2_{r-s}(\mu'M(M'\Sigma M)^{-1}M'\mu).$$

Hieraus folgt wegen $\hat{M}_n \xrightarrow{\mathbb{P}_n} M^+$, $\hat{\Sigma}_n \xrightarrow{\mathbb{P}_n} \Sigma$ die Beziehung

$$\sqrt{S_n} (\hat{M}'_n \hat{\Sigma}_n \hat{M}_n)^{-1/2} \hat{R}_n - \sqrt{S_n} (M'\Sigma M)^{-1/2} \hat{R}_n \xrightarrow{\mathbb{P}_n} 0$$

und somit

$$(7.7) \quad W_n - S_n \hat{R}'_n (M'\Sigma M)^{-1} \hat{R}_n \xrightarrow{\mathbb{P}_n} 0.$$

Aus (7.6) und (7.7) ergibt sich die Behauptung. □

Satz 7.2

Gelten dieselben Voraussetzungen wie bei Satz 7.1. Existiere außerdem eine Konstante $c > 0$, so daß die wie im vorhergehenden Satz definierten Matrizen Σ , A und J der Bedingung $A'\Sigma A = cJ$ genügen. Dann sind die durch (7.1) - (7.3) definierten Statistiken W_n , C_n und $2\Lambda_n/c$ asymptotisch äquivalent, d.h.

$$(7.8) \quad W_n - C_n \xrightarrow{\mathbb{P}_n} 0$$

und

$$(7.9) \quad W_n - 2\Lambda_n/c \xrightarrow{\mathbb{P}_n} 0.$$

Außerdem sind sie asymptotisch nichtzentral χ^2 -verteilt mit $r-s$ Freiheitsgraden und dem Nichtzentralitätsparameter

$$(7.10) \quad \begin{aligned} \delta &= \mu'A(A'\Sigma A)^{-1}T'[T(A'\Sigma A)^{-1}T']^{-1}T(A'\Sigma A)^{-1}A'\mu \\ &= \mu'[A(A'\Sigma A)^{-1}A' - AF(F'A'\Sigma AF)^{-1}F'A']\mu, \end{aligned}$$

+) Ist eine Variable X_n nicht überall definiert, so ist $X_n \xrightarrow{\mathbb{P}_n} X$ zu lesen als $X_n^* \xrightarrow{\mathbb{P}_n} X$ mit $X_n^* := \begin{cases} X_n, & \text{falls } X_n \text{ definiert ist} \\ \text{beliebig,} & \text{sonst} \end{cases}$

wobei $T := DR(\theta_0)$ und $F := Df(\beta_0)$ mit $\beta_0 := f^{-1}(\theta_0) = (f \circ f)^{-1}(\theta_0)$ gelten und $f(\cdot)$ wie in (c) definiert ist. Im Fall $s = 0$ reduziert sich (7.10) zu

$$(7.11) \quad \delta = \mu' A (A' \Sigma A)^{-1} A' \mu.$$

Sind zusätzlich

$$(7.12) \quad \begin{cases} h_i(a, a) = D_1 h_i(a, a) = 0, \\ D_1^2 h_i(a, a) = D_2 g_i(a, a) = -D_1 g_i(a, a) \quad \forall a \in (0, 1), i=1, \dots, k, \end{cases}$$

erfüllt, so gilt außerdem

$$(7.13) \quad 2 \mathcal{D}(\hat{\pi}_n, \hat{\pi}_n^{\#}, N_n) / c - W_n \xrightarrow{P_n} 0.$$

Bemerkung:

Unter der zusätzlichen Annahme (6.20) lassen sich (7.10) und (7.11) äquivalent schreiben als

$$\delta = [(P_G - P_{GF})\mu]' \Sigma^{-1} (P_G - P_{GF})\mu$$

und

$$\delta = (P_G \mu)' \Sigma^{-1} P_G \mu,$$

wobei $G := D\pi(\theta_0)$ gelte und P_G bzw. P_{GF} die orthogonale Projektion auf den Tangentialraum $\mathcal{M}(G)$ durch π_0 an $\mathcal{J}$ bzw. auf den Tangentialraum $\mathcal{M}(GF)$ durch π_0 an $\mathcal{J}'$ bzgl. des Skalarprodukts $\langle X, Y \rangle := Y' \Sigma^{-1} X$ ist.

Beweis:

Zunächst beweisen wir die zweite Gleichung in (7.10). Sei $B := A' \Sigma A$ und sei $\mathcal{M}(T')$ der von den Zeilen von T aufgespannte lineare Teilraum im $\mathbb{R}^r$. Dann ist $P_{T'} := T'(TB^{-1}T')^{-1}TB^{-1}$ eine orthogonale Projektion auf $\mathcal{M}(T')$ bzgl. des Skalarprodukts $\langle X, Y \rangle := Y'B^{-1}X$ (Rao (73), S. 47). Entsprechend ist die orthogonale Projektion bzgl. desselben Skalarprodukts auf $\mathcal{M}(BF)$ gegeben durch $P_{BF} := BF(F'B F)^{-1}F'$. Wegen $TF = DR(\theta_0) Df(\beta_0) = D(R \circ f)(\beta_0) = 0$ gilt $P_{T'} + P_{BF} = I$ und daher

$$(7.14) \quad T'(TB^{-1}T')^{-1}TB^{-1} = I - BF(F'BF)^{-1}F'.$$

Hieraus folgt die zweite Gleichung in (7.10).

Eine Taylorentwicklung und Anwendung von Satz 6.3 liefert

$$\begin{aligned} \sqrt{S_n} \hat{R}_n &= \sqrt{S_n} T (\hat{\theta}_n - \theta_0) + o_{P_n}(1) \\ &= -TJ^{-1}A'\sqrt{S_n} (P_n - \pi_0) + o_{P_n}(1) \end{aligned}$$

und daher unter Beachtung von $B = A'\Sigma A = cJ$, (7.7) und (7.14)

$$\begin{aligned} (7.15) \quad W_n &= U_n' AB^{-1} T'(TB^{-1}T')^{-1} TB^{-1} A' U_n + o_{P_n}(1) \\ &= U_n' A [B^{-1} - F(F'BF)^{-1}F'] A' U_n + o_{P_n}(1), \end{aligned}$$

wobei $U_n := \sqrt{S_n} (P_n - \pi_0)$ gelte.

Weiter liefert die Taylorentwicklung von $g_i(\cdot, \pi_i(\cdot))$ um $(\pi_{0i}, \pi_i(\theta_0))$, $i=1, \dots, k$, die Beziehung

$$\begin{aligned} g_i(P_{ni}, \hat{\pi}_{ni}) &= [D_1 g_i(\pi_{0i}, \pi_{0i}) + o_{P_n}(1)] (P_{ni} - \pi_{0i}) + \\ &+ \sum_{j=1}^r [D_2 g_i(\pi_{0i}, \pi_{0i}) \frac{\partial \pi_i}{\partial \theta_j}(\theta_0) + o_{P_n}(1)] (\hat{\theta}_{nj} - \theta_0) \\ &\quad i=1, \dots, k. \end{aligned}$$

Nach Multiplikation beider Seiten mit $N_{ni}/\sqrt{S_n} \frac{\partial \pi_i}{\partial \theta_t}(\hat{\theta}_n)$, $t=1, \dots, r$ und ähnlichem Vorgehen wie im Beweis zu Satz 6.3 erhalten wir in Matrixschreibweise

$$(7.16) \quad \sqrt{S_n} \hat{Z}_n = A' U_n + \sqrt{S_n} J (\hat{\theta}_n - \theta_0) + o_{P_n}(1).$$

Da aufgrund der Taylorentwicklung von $f(\cdot)$ um β_0

$$(7.17) \quad \sqrt{S_n} (\hat{\theta}_n - \theta_0) = \sqrt{S_n} F (\hat{\beta}_n - \beta_0) + o_{P_n}(1),$$

$\hat{\beta}_n := f^{-1}(\hat{\theta}_n)$, und nach Satz 6.3

$$(7.18) \quad \sqrt{S_n} (\hat{\beta}_n - \beta_0) = -(F'JF)^{-1}F'A'U_n + o_{P_n}(1)$$

gelten, bekommen wir aus (7.16) unter Beachtung von $B = cJ$

$$\sqrt{S_n} \hat{\hat{Z}}_n = A' U_n - BF(F'BF)^{-1} F' A' U_n + o_{P_n}(1).$$

Mit (7.15) folgt hieraus

$$\begin{aligned} S_n \hat{\hat{Z}}_n' B^{-1} \hat{\hat{Z}}_n &= U_n' A [B^{-1} - F(F'BF)^{-1} F'] A' U_n + o_{P_n}(1) \\ &= W_n + o_{P_n}(1), \end{aligned}$$

woraus wir mit ähnlicher Argumentation wie bei der Herleitung von (7.7)

$$C_n = W_n + o_{P_n}(1)$$

erhalten. Damit ist (7.8) bewiesen.

Weiter liefert die Taylorentwicklung von $h_i(\cdot)$ um (π_{oi}, π_{oi}) , $i=1, \dots, k$, unter Benutzung von $g_i(\pi_{oi}, \pi_{oi}) = 0$ und

$$S_n (P_{ni} - \pi_{oi})^2 = o_{P_n}(1), \quad S_n (\hat{\pi}_{ni} - \pi_{oi})^2 = o_{P_n}(1) \quad \text{für } i=1, \dots, k$$

$$\begin{aligned} h_i(P_{ni}, \pi_{ni}) &= h_i(\pi_{oi}, \pi_{oi}) + D_1 h_i(\pi_{oi}, \pi_{oi}) (P_{ni} - \pi_{oi}) + \\ &+ \frac{1}{2} D_2 g_i(\pi_{oi}, \pi_{oi}) (\hat{\pi}_{ni} - \pi_{oi})^2 + \\ &+ D_1 g_i(\pi_{oi}, \pi_{oi}) (\hat{\pi}_{ni} - \pi_{oi}) (P_{ni} - \pi_{oi}) + \\ &+ \frac{1}{2} D_1^2 h_i(\pi_{oi}, \pi_{oi}) (P_{ni} - \pi_{oi})^2 + \frac{1}{S_n} o_{P_n}(1). \end{aligned}$$

Eine analoge Entwicklung läßt sich für $h_i(P_{ni}, \hat{\pi}_{ni})$ angeben. Wir erhalten dann unter Benutzung von (7.17), (7.18) und Satz 6.3

$$2\Lambda_n/c = 2 \sum_{i=1}^k N_{ni} [h_i(P_{ni}, \hat{\pi}_{ni}) - h_i(P_{ni}, \hat{\pi}_{ni})] / c$$

$$\begin{aligned}
 &= \sum_{i=1}^k N_{ni} [D_2 g_i(\pi_{oi}, \pi_{oi}) (\hat{\pi}_{ni} - \pi_{oi})^2 + \\
 &\quad + 2 D_1 g_i(\pi_{oi}, \pi_{oi}) (P_{ni} - \pi_{oi}) (\hat{\pi}_{ni} - \pi_{oi}) - \\
 &\quad - D_2 g_i(\pi_{oi}, \pi_{oi}) (\hat{\pi}_{ni} - \pi_{oi})^2 - 2 D_1 g_i(\pi_{oi}, \pi_{oi}) \cdot \\
 &\quad \cdot (P_{ni} - \pi_{oi}) (\hat{\pi}_{ni} - \pi_{oi})] / c + o_{P_n}(1) \\
 &= [S'_n (\hat{\theta}_n - \theta_0)' J (\hat{\theta}_n - \theta_0) + 2 \sqrt{S}_n (\hat{\theta}_n - \theta_0)' A' U_n - \\
 &\quad - S_n (\hat{\theta}_n - \theta_0)' J (\hat{\theta}_n - \theta_0) - 2 \sqrt{S}_n (\hat{\theta}_n - \theta_0)' A' U_n] / c + o_{P_n}(1) \\
 &= U'_n A F (F' B F)^{-1} F' A' U_n - 2 U'_n A F (F' B F)^{-1} F' A' U_n - \\
 &\quad - U'_n A B^{-1} A' U_n + 2 U'_n A B^{-1} A' U_n + o_{P_n}(1) \\
 &= U'_n A [B^{-1} - F (F' B F)^{-1} F'] A' U_n + o_{P_n}(1),
 \end{aligned}$$

wobei bei der vorletzten Gleichung die Beziehung $B = cJ$ benutzt wurde. Wegen (7.15) gilt deshalb

$$2\Lambda_n / c = W_n + o_{P_n}(1),$$

womit (7.9) bewiesen ist. Daß W_n - und damit auch C_n und $2\Lambda_n / c$ - asymptotisch $\chi^2_{r-s}(\delta)$ -verteilt sind, wobei δ durch (7.10) gegeben ist, folgt sofort aus Satz 7.1.

(7.13) sieht man, wenn man $h_i(\cdot)$ um (π_{oi}, π_{oi}) , $i=1, \dots, k$, nach Taylor entwickelt und $g_i(\pi_{oi}, \pi_{oi}) = 0$ und die Zusatzbedingungen (7.12) einsetzt:

$$\begin{aligned}
 h_i(\hat{\pi}_{ni}, \hat{\pi}_{ni}) &= \frac{1}{2} D_2 g_i(\pi_{oi}, \pi_{oi}) (\hat{\pi}_{ni} - \pi_{oi})^2 + \\
 &\quad + \frac{1}{2} D_2 g_i(\pi_{oi}, \pi_{oi}) (\hat{\pi}_{ni} - \pi_{oi})^2 - \\
 &\quad - D_2 g_i(\pi_{oi}, \pi_{oi}) (\hat{\pi}_{ni} - \pi_{oi}) (\hat{\pi}_{ni} - \pi_{oi}) + \\
 &\quad + o_{P_n}(1), \quad i=1, \dots, k.
 \end{aligned}$$

Hieraus folgt analog zu dem vorhergehenden Beweis

$$\begin{aligned}
 2\mathcal{D}(\hat{\pi}_n, \hat{\pi}_n, N_n) / c &= [S_n(\hat{\theta}_n - \theta_0)' J(\hat{\theta}_n - \theta_0) + S_n(\hat{\hat{\theta}}_n - \theta_0)' J(\hat{\hat{\theta}}_n - \theta_0) - \\
 &\quad - 2 S_n(\hat{\theta}_n - \theta_0)' J(\hat{\hat{\theta}}_n - \theta_0)] / c + o_{P_n}(1) \\
 &= U_n' A B^{-1} A' U_n + U_n' A F(F' B F)^{-1} F' A' U_n - \\
 &\quad - 2 U_n' A F(F' B F)^{-1} F' A' U_n + o_{P_n}(1) \\
 &= U_n' A [B^{-1} - F(F' B F)^{-1} F'] A' U_n + o_{P_n}(1) \\
 &= W_n + o_{P_n}(1).
 \end{aligned}$$

Damit ist (7.13) bewiesen. □

Die Bedingung $A' \Sigma A = cJ$ und die Zusatzbedingungen (7.12) sind bei dem Maximum-Likelihood-Schätzer und bei den beiden Minimum- χ^2 -Schätzern erfüllt.

Die Sätze 7.1 und 7.2 bleiben gültig, wenn bei den Teststatistiken (7.1), (7.2) oder (7.3) der Schätzer $\hat{\theta}_n$ bzw. $\hat{\hat{\theta}}_n$ durch einen asymptotisch äquivalenten Schätzer $\tilde{\theta}_n$ bzw. $\tilde{\hat{\theta}}_n$ ersetzt wird, d.h.

$$\sqrt{S_n}(\hat{\theta}_n - \tilde{\theta}_n) \xrightarrow{P_n} 0$$

bzw.

$$\sqrt{S_n}(\hat{\hat{\theta}}_n - \tilde{\hat{\theta}}_n) \xrightarrow{P_n} 0$$

gilt. Da der Maximum-Likelihood-Schätzer und die beiden Minimum- χ^2 -Schätzer asymptotisch äquivalent sind, kann man z.B. für $\mathcal{D}(\cdot)$ eine der χ^2 -Abstandsfunktion zugrundelegen, während man für $\hat{\theta}_n$ und $\hat{\hat{\theta}}_n$ bzw. $\hat{\pi}_n$ und $\hat{\hat{\pi}}_n$ den Maximum-Likelihood-Schätzer einsetzt.

II. KONTINGENZTAFELN MIT UNENDLICH VIELEN ZELLEN

8. Modell

Entsprechend den Ausführungen in der Einleitung legen wir folgendes Modell zugrunde, welches wir gleich für den infiniten Fall und unter Einschluß einiger Regularitätsvoraussetzungen formulieren:

Sei $(\Omega, \mathcal{A}, \mathbb{P}_\theta)$ für jedes $\theta \in \Theta \subset \mathbb{R}^r$, $r > 0$, ein Wahrscheinlichkeitsraum und $\{N_n, D_n, X_n\}$ eine Folge von unabhängigen Zufallsvektoren auf $(\Omega, \mathcal{A})$, wobei für alle $n \in \mathbb{N}$ $N_n = (N_{n1}, \dots, N_{nk})$ und $D_n = (D_{n1}, \dots, D_{nk})$, $k > 0$, Werte aus $\{0, \dots, M\}^k$, $M > 0$, und (die Kovariablen) $X_n = (X_{n1}, \dots, X_{nt})$, $t > 0$, Werte aus $\mathbb{X} \subset \mathbb{R}^t$ annehmen.

Weiter existiere für jedes $\theta \in \Theta$ eine Familie $\{\mathbb{P}_\theta^x: x \in \mathbb{X}\}$ von Wahrscheinlichkeitsmaßen auf $(\Omega, \mathcal{A})$ derart, daß für alle $m, d \in \{0, \dots, M\}^k$ und $n \in \mathbb{N}$ $\mathbb{P}_\theta^x\{N_n \leq m, D_n \leq d\}$ stetig in $x \in \mathbb{X}$ ist; $\mathbb{P}_\theta\{N_n \leq m, D_n \leq d | X_n\} = \mathbb{P}_0^x\{N_n \leq m, D_n \leq d\}$ $\mathbb{P}_\theta$ -f.s. gilt und die folgenden Bedingungen für jedes $x \in \mathbb{X}$ erfüllt sind:

I. Die Verteilung von (N_n, D_n) bzgl. $\mathbb{P}_\theta^x$ hängt nicht von n ab:

$$\mathbb{P}_\theta^x\{N_n \leq m, D_n \leq d\} = \mathbb{P}_\theta^x\{N_1 \leq m, D_1 \leq d\} \quad \forall m, d \in \{0, \dots, M\}^k, n=2,3,\dots$$

II. Für alle $n \in \mathbb{N}$, $i=1, \dots, k$ ist D_{ni} unter der Bedingung N_{ni} bzgl. $\mathbb{P}_\theta^x$ binomialverteilt:

$$\mathcal{L}_\theta^x(D_{ni} | N_{ni}=m) = \mathcal{B}(m, \rho_i(x, \theta)) \quad \forall m \in \{0, \dots, M\} \text{ mit } \mathbb{P}_\theta^x\{N_{ni}=m\} > 0,$$

wobei $\rho_i(\cdot)$ eine stetige Funktion $\mathbb{X} \times \Theta \rightarrow (0, 1)$ ist.

III. Für alle $n \in \mathbb{N}$, $i=1, \dots, k$ und $d \in \{0, \dots, M\}$ gilt

$$\mathbb{P}_\theta^x\{D_{ni}=d | N_{n1}, D_{n1}, \dots, N_{ni}\} = \mathbb{P}_\theta^x\{D_{ni}=d | N_{ni}\} \quad \mathbb{P}_\theta^x\text{-f.s.}$$

IV. Für alle $n \in \mathbb{N}$, $i=1, \dots, k$ gilt

$$\mathbb{P}_\theta^x\{N_{ni} > 0\} > 0.$$

Für jedes $\theta \in \Theta$ gelte außerdem die folgende Bedingung über die

Art der Asymptotik:

V. Die Folge $\{\hat{G}_n(\cdot)\}$ der empirischen Verteilungsfunktionen von $X_1, \dots, X_n$, $n \in \mathbb{N}$, definiert durch

$$(8.1) \quad \hat{G}_n(x) := \frac{1}{n} \sum_{l=1}^n I_{(-\infty, x]}(X_l) \quad \forall x \in \mathbb{R}^t$$

konvergiert $\mathbb{P}_\theta$ -f.s. gegen eine Verteilungsfunktion $G_\theta(\cdot)$:

$$(8.2) \quad \sup_{x \in \mathbb{R}^t} \|\hat{G}_n(x) - G_\theta(x)\| \xrightarrow{\mathbb{P}_\theta\text{-f.s.}} 0.$$

Im folgenden sei

$$(8.3) \quad H_\theta(m, d; x) := \mathbb{P}_\theta^x\{N_1 \leq m, D_1 \leq d\} \quad \forall m, d \in \{0, \dots, \}, x \in \mathcal{X}, \theta \in \Theta.$$

Außerdem definieren wir für jede Zufallsvariable Z_n der Form $Z_n = f(N_n, D_n, X_n)$, $n \in \mathbb{N}$, ($f(\cdot)$ meßbar)

$$(8.4) \quad \begin{aligned} \mathbb{E}_\theta\{Z_n; x\} &:= \int f(N_n, D_n, x) d\mathbb{P}_\theta^x \\ &= \int f(m, d, x) H_\theta(d(m, d); x) \quad \forall x \in \mathcal{X}, \theta \in \Theta. \end{aligned}$$

Existiert (für festes $x \in \mathcal{X}$) $\mathbb{E}_\theta\{Z_n; x\}$ und gilt $\mathbb{E}_\theta\{|Z_n|\} < \infty$, so ist $\mathbb{E}_\theta\{Z_n; x\}$ eine Version des bedingten Erwartungswertes $\mathbb{E}_\theta\{Z_n | X_n = x\}$ von Z_n gegeben $X_n = x$ (Gänssler, Stute(77), S.195). Natürlich gilt

$$\mathbb{E}_\theta\{Z_n; x\} = \mathbb{E}_\theta\{Z_1; x\} \quad \forall x \in \mathcal{X}, \theta \in \Theta, n=2,3,\dots$$

Die Bedingung (8.2) ist nach der multivariaten Version des Satzes von Glivenko - Cantelli (Gänssler, Stute (77), S.145) z.B. dann erfüllt, wenn $\{X_n\}$ eine Folge von identisch und unabhängig verteilten Zufallsvariablen ist. Falls die X_n Einpunktverteilungen besitzen und der Kovariablenraum $\mathcal{X}$ endlich ist - etwa $\mathcal{X} = \{x_1, \dots, x_s\}$, so ist die Bedingung erfüllt, wenn die relativen Häufigkeiten $\frac{1}{n} \sum_{l=1}^n I_{\{x_i\}}(X_l)$, $i=1, \dots, s$, konvergieren. Dieser Fall ist von Bedeutung, wenn die Kovariablen vom Experimentator gesetzt werden. $G_\theta(\cdot)$ wird dann zweckmäßigerweise als empirische Verteilungsfunktion der im Experiment gesetzten Kovariablenwerte gewählt. Die Bedingung (8.2) läßt sich natürlich entsprechend erfüllen, wenn einige Komponenten der Kovariablen gesetzt werden und die übrigen für jede gegebene Konstellation der gesetzten

Kovariablen identisch und unabhängig verteilt sind.

Bei Anwendungen ist es Aufgabe, aufgrund von Beobachtungen $(N_1, D_1, X_1), \dots, (N_n, D_n, X_n)$ den zugrundeliegenden unbekannten Parameter aus Θ zu schätzen. Wir bezeichnen diesen Parameter im folgenden mit θ_0 . Ziel dieses Kapitels ist es, für eine Klasse von Schätzern (unter einigen Regularitätsbedingungen) Konsistenz bzgl. θ_0 und asymptotische Normalität unter dem (im Gegensatz zum ersten Kapitel festen) Wahrscheinlichkeitsmaß P_{θ_0} herzuleiten.

9. Beispiele

9.1 Quantal - Response - Modell (Binomialmodell)

Sei $\{X_n, R_n\}$ eine Folge von unabhängigen Zufallsvariablen, wobei die Kovariablen X_n die Modellbedingung V erfüllen und die R_n bedingt auf X_n binomialverteilt sind:

$$\mathcal{L}_{\theta}^X(R_n) = \mathcal{B}(1, \rho(x, \theta)) \quad \forall x \in \mathcal{X} \subset \mathbb{R}^t, \theta \in \Theta \subset \mathbb{R}^r,$$

wobei $\rho(\cdot)$ eine stetige Funktion $\mathcal{X} \times \Theta \rightarrow (0, 1)$ ist. Dieses Beispiel läßt sich dem Modell unterordnen, indem $k = 1$, $D_{n1} = R_n$ und $N_{n1} \equiv 1 \quad \forall n \in \mathbb{N}$ gesetzt werden.

9.2 Multinomialmodell

Sei $\{X_n, D_{n1}, \dots, D_{n, k+1}\}$ eine Folge von unabhängigen Zufallsvariablen, wobei die Kovariablen X_n die Modellbedingung V erfüllen und die $(D_{n1}, \dots, D_{n, k+1})$ bedingt auf X_n multinomialverteilt sind:

$$\mathcal{L}_{\theta}^X(D_{n1}, \dots, D_{n, k+1}) = \mathcal{M}(1, P(x, \theta)) \quad \forall x \in \mathcal{X} \subset \mathbb{R}^t, \theta \in \Theta \subset \mathbb{R}^r,$$

wobei $P(\cdot) = (P_1(\cdot), \dots, P_{k+1}(\cdot))$ eine stetige Funktion $\mathcal{X} \times \Theta \rightarrow (0, 1)^{k+1}$ mit $\sum_{i=1}^{k+1} P_i(\cdot) \equiv 1$ ist. Dieses Beispiel läßt sich dem Modell unterordnen, indem man für jedes $n \in \mathbb{N}$

$$N_{ni} := 1 - \sum_{j=1}^{i-1} D_{nj}, \quad i=1, \dots, k,$$

und

$$\rho_i(\cdot) := P_i(\cdot) / \sum_{j=1}^{k+1} P_j(\cdot), \quad i=1, \dots, k,$$

setzt.

Haberman (74, S. 352ff) leitet die asymptotische Normalität des Maximum-Likelihood-Schätzers für den speziellen Fall her, daß die X_n Einpunktverteilungen besitzen und $P_i(\cdot)$, $i=1, \dots, k+1$, durch

$$P_i(x, \theta) := \frac{e^{\beta'_i x}}{\sum_{j=1}^{k+1} e^{\beta'_j x}} \quad \forall x \in \mathcal{X}, \theta \in \Theta,$$

mit $\theta' := (\beta'_1, \dots, \beta'_k)$, $\sum_{j=1}^{k+1} \beta_j = 0$, gegeben ist (multivariates Logitmodell).

9.3 Diskretes Lebenszeitmodell

Dieses Modell wurde bereits ausführlich in der Einleitung beschrieben. Im speziellen Diskreten - Proportionalen - Hazards - Modell (siehe Einleitung) gilt

$$\rho_i(x, \theta) = F(\gamma_i + \beta'_i x) \quad , i=1, \dots, k,$$

wobei $\theta' = (\gamma_1, \dots, \gamma_k, \beta')$, $\gamma_i := \log(-\log(1-\lambda_i))$, $i=1, \dots, k$, und $F(u) := 1 - \exp(-e^u)$, $u \in \mathbb{R}$, (komplementäre log-log - Verteilung) gilt.

10. Schätzmethoden

Wir schätzen den unbekannten Parameter θ wie in Kapitel I bzgl. einer Funktion $\mathcal{D}(\cdot): [0,1]^k \times (0,1)^k \times [0,\infty)^k \rightarrow [0,\infty)$, welche die Form

$$\mathcal{D}(p, \tau, q) = \sum_{i=1}^k q_i h_i(p_i, \tau_i)$$

besitzt. Im folgenden seien $\rho(\cdot) := (\rho_1(\cdot), \dots, \rho_k(\cdot))'$ und $P_n := (P_{n1}, \dots, P_{nk})'$, $n \in \mathbb{N}$, mit

$$(10.1) \quad P_{ni} := \begin{cases} D_{ni}/N_{ni}, & \text{falls } N_{ni} > 0, D_{ni} \leq N_{ni}, \\ 0, & \text{sonst} \end{cases} \quad i=1, \dots, k.$$

Wir definieren den Schätzer $\hat{\theta}_n$ bzgl. $\mathcal{D}(\cdot)$ als irgendeine (nicht notwendig meßbare) Variable $\Omega \rightarrow \Theta$, die die Gleichung

$$(10.2) \quad \sum_{l=1}^n \mathcal{D}(P_l, \rho(X_l, \hat{\theta}_n), N_l) = \inf_{\theta \in \Theta} \sum_{l=1}^n \mathcal{D}(P_l, \rho(X_l, \theta), N_l)$$

erfüllt, sofern das Infimum für ein $\theta \in \Theta$ angenommen wird.

Wir diskutieren im folgenden insbesondere den Maximum-Likelihood-Schätzer (definiert durch (4.7)), den Minimum- χ^2 -Schätzer (definiert durch (4.8)) und den Einfachen Schätzer (definiert durch (4.10)). Der verallgemeinerte modifizierte Minimum- χ^2 -Schätzer, definiert durch (4.9), ist nicht sinnvoll, da jedes D_{ni} mit positiver Wahrscheinlichkeit die Werte 0 oder N_{ni} annehmen kann und hierfür die Abstandsfunktion (4.9) nicht spezifiziert ist. Es wird sich herausstellen, daß der Minimum- χ^2 -Schätzer im allgemeinen nicht konsistent ist, während die beiden anderen (unter geeigneten Regularitätsannahmen) sowohl konsistent als auch asymptotisch normalverteilt sind.

11. Konsistenz

Konsistenz und asymptotische Normalität werden zunächst für ein allgemeineres Modell gezeigt, da es (im Gegensatz zum Modell in Kapitel I) keine Vorteile bringt, sich auf das spezielle Kontingenztafelmodell zu beschränken und sich die Regularitätsbedingungen allgemein eleganter formulieren lassen. Anschließend prüfen wir die Regularitätsbedingungen für unser spezielles Modell.

Allgemeines Modell:

Sei $(\Omega, \mathcal{A}, \mathbb{P}_\theta)$ für jedes $\theta \in \Theta \subset \mathbb{R}^r$, $r > 0$, ein Wahrscheinlichkeitsraum und $\{Y_n\}$ mit $Y_n := (W_n, X_n)$ eine Folge von unabhängigen Zufallsvariablen auf $(\Omega, \mathcal{A})$ mit Werten in einem Meßraum $(W \times X, \mathcal{B}^{s-t} \times \mathcal{B}^t)$, $W \subset \mathbb{R}^{s-t}$ abgeschlossen, $X \subset \mathbb{R}^t$ ($0 < t < s$), und den folgenden Eigenschaften für jedes $\theta \in \Theta$:

- es existiert für alle $x \in X$ eine Verteilungsfunktion $H_\theta(\cdot; x)$ auf $\mathbb{R}^{s-t}$, so daß

$$(11.1) \quad \mathbb{P}_\theta\{W_n \leq w | X_n\} = H_\theta(w; X_n) \quad \mathbb{P}_\theta\text{-f.s.} \quad \forall w \in W, n \in \mathbb{N}$$

gilt und $H_\theta(w; \cdot)$ für alle $w \in \mathbb{R}^{s-t}$ stetig ist,

- es existiert eine Verteilungsfunktion $G_\theta(\cdot)$ derart, daß

$$(11.2) \quad \sup_{x \in \mathbb{R}^t} \|\hat{G}_n(x) - G_\theta(x)\| \xrightarrow{\mathbb{P}_\theta\text{-f.s.}} 0$$

gilt, wobei $\hat{G}_n(\cdot)$ die empirische Verteilungsfunktion von $X_1, \dots, X_n$, definiert durch (8.1), ist.

Im Allgemeinen Modell sei der Schätzer $\hat{\theta}_n$ bzgl. einer vorgegebenen nichtnegativen Funktion $\phi(\cdot): \mathcal{Y} \times \Theta \rightarrow [0, \infty)$ mit $J = W \times X$ als irgendeine Variable $J \rightarrow \Theta$ definiert mit

$$(11.3) \quad \sum_{l=1}^n \phi(Y_l, \hat{\theta}_n) = \inf_{\theta \in \Theta} \sum_{l=1}^n \phi(Y_l, \theta),$$

sofern das Infimum für ein $\theta \in \Theta$ angenommen wird. (Ein derart definierter Schätzer wird auch "M-Schätzer" genannt.)

Wir werden Konsistenz und asymptotische Normalität unter einem festen Wahrscheinlichkeitsmaß P_{θ_0} zeigen, wobei θ_0 irgendein Punkt aus Θ ist. Der Einfachheit halber schreiben wir im folgenden P statt P_{θ_0} , E statt E_{θ_0} , G statt G_{θ_0} usw..

Im weiteren sei $Y = (W, X)$ eine Zufallsvariable mit Werten in $(W \times X, \mathcal{B}^S)$, welche die Verteilungsfunktion

$$Q(w, x) := \int_{(-\infty, x]} H(w; u) G(du), \quad w \in \mathbb{R}^{s-t}, \quad x \in \mathbb{R}^t,$$

besitzt. (Daß $Q(\cdot)$ eine Verteilungsfunktion ist, sieht man wie folgt: Sei für alle $x \in X$ $\mu(\cdot; x)$ das zu $H(\cdot; x)$ gehörende Wahrscheinlichkeitsmaß. Dann ist $\mu(A; \cdot)$ für alle A aus der Klasse $\{(-\infty, a] : a \in \mathbb{R}^{s-t}\}$ meßbar. Da diese Klasse $\mathcal{B}^{s-t}$ erzeugt und $\cap$ -stabil ist, folgt daraus die Meßbarkeit von $\mu(A; \cdot)$ für alle $A \in \mathcal{B}^{s-t}$ (Gänssler, Stute (77), S.43). Damit sind die Bedingungen erfüllt, damit durch

$$\nu(A_1 \times A_2) := \int_{A_2} \mu(A_1; x) G(dx), \quad A_1 \in \mathcal{B}^{s-t}, \quad A_2 \in \mathcal{B}^t$$

ein eindeutiges Wahrscheinlichkeitsmaß auf $(W \times X, \mathcal{B}^S)$ bestimmt wird (Loève (77a), S.25).)

Insbesondere besitzt $Y = (W, X)$ damit die Eigenschaften

$$P\{X \leq x\} = G(x) \quad \forall x \in X$$

und

$$P\{W \leq w | X\} = H(w; X) \quad P\text{-f.s.} \quad \forall w \in W$$

(, wobei hier P ein Wahrscheinlichkeitsmaß auf einem geeigneten Meßraum ist, der im allgemeinen von $(\Omega, \mathcal{A})$ verschieden ist).

Entsprechend (8.4) definieren wir für jede meßbare Funktion $f(\cdot): Y \rightarrow \mathbb{R}$

$$(11.4) \quad E\{f(Y_1); x\} := \int f(w, x) H(dw; x) \quad \forall x \in X.$$

Existiert für jedes $x \in X$ $E\{f(Y_1); x\}$ und ist $E\{f(Y)\}$ definiert, so gilt insbesondere

$$(11.5) \quad \int_X E\{f(Y_1); x\} G(dx) = E\{f(Y)\}$$

(Gänssler, Stute (77), S. 45).

Unsere Herleitung ist anders als die von Hoadley (71) gegebene. Er beweist das asymptotische Verhalten des Maximum-Likelihood-Schätzers im unabhängigen, nicht identisch verteilten Fall ohne Einführung von Kovariablen. Durch die Formulierung mit Kovariablen werden meiner Meinung nach sowohl die Regularitätsbedingungen als auch die Beweise anschaulicher. Der zentrale Punkt unseres Ansatzes ist die Einführung der Zufallsvariablen Y , wodurch es uns ermöglicht wird, weitgehend auf die herkömmliche Methode für den Fall zurückzugreifen, das $Y_1, Y_2, \dots$ unabhängig und identisch verteilt sind (Konsistenz: Wald (49); asymptotische Normalität: Witting, Nölle (70), S.78). Ganz grob gesprochen, besteht die Methode im unabhängig und identisch verteilten Fall aus drei Schritten:

(i) Herleitung gewisser Eigenschaften von Y_1 .

(ii) Anwendung des Gesetzes der großen Zahlen:

$$\frac{1}{n} \sum_{l=1}^n f(Y_l) \xrightarrow{\text{P-f.s.}} \quad (\text{bzw. } \xrightarrow{\text{P}}) \quad E\{f(Y_1)\}$$

für gewisse Funktionen $f(\cdot)$.

(iii) Anwendung des zentralen Grenzwertsatzes (nur für asymptotische Normalität).

Unsere Vorgehensweise unterscheidet sich von dieser Methode nur dadurch, daß in (i) Y die Rolle von Y_1 und in (ii) $E\{f(Y)\}$ die Rolle von $E\{f(Y_1)\}$ übernimmt.

Unser erstes Ziel ist zu zeigen, daß für jede stetige Funktion $f(\cdot): Y \rightarrow \mathbb{R}$ mit $\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{l=1}^n E\{|f(Y_l)|^{1+\delta}\} < \infty$ für ein $\delta > 0$ der Erwartungswert $E\{f(Y)\}$ existiert und die Beziehung

$$(11.6) \quad \frac{1}{n} \sum_{l=1}^n f(Y_l) \xrightarrow{\text{P-f.s.}} E\{f(Y)\}$$

erfüllt ist. Hierfür beweisen wir zunächst zwei Lemmata:

Lemma 11.1

Sei $\{a_n\}$ eine Folge nichtnegativer Zahlen. Dann folgt aus

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{l=1}^n a_l < \infty$$

die Beziehung

$$\sum_{n=1}^{\infty} \frac{1}{n^{1+\delta}} a_n < \infty \quad \forall \delta > 0.$$

Beweis:

Nach Voraussetzung existiert eine Schranke $K > 0$, so daß

$$\sum_{l=1}^n a_l \leq nK \quad \forall n \in \mathbb{N}.$$

Seien nun $\delta > 0$ beliebig und $b_n := a_n - K \quad \forall n \in \mathbb{N}$. Dann gilt

$$\sum_{l=1}^n \frac{1}{l^{1+\delta}} a_l = K \sum_{l=1}^n \frac{1}{l^{1+\delta}} + \sum_{l=1}^n \frac{1}{l^{1+\delta}} b_l \quad \forall n \in \mathbb{N}.$$

Definiere nun für festes $n \in \mathbb{N}$ $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ durch

$$\frac{1}{l^{1+\delta}} = \sum_{j=1}^{n-l+1} \alpha_j, \quad l=1, \dots, n.$$

Dann gilt wegen $\alpha_j > 0$, $j=1, \dots, n$, und wegen $\sum_{l=1}^m b_l \leq 0$, $m=1, \dots, n$, die Beziehung

$$\sum_{l=1}^n \frac{1}{l^{1+\delta}} b_l = \sum_{l=1}^n \sum_{j=1}^{n-l+1} \alpha_j b_l = \sum_{j=1}^n \alpha_j \sum_{l=1}^{n-j+1} b_l \leq 0.$$

Da $n \in \mathbb{N}$ beliebig war, erhalten wir somit

$$\sum_{l=1}^n \frac{1}{l^{1+\delta}} a_l \leq K \sum_{l=1}^n \frac{1}{l^{1+\delta}} \quad \forall n \in \mathbb{N}.$$

Hieraus ergibt sich wegen $\sum_{l=1}^{\infty} \frac{1}{l^{1+\delta}} < \infty$ die Behauptung. □

Lemma 11.2

Sei $f(\cdot)$ eine stetige Funktion $A \times K \rightarrow \mathbb{R}$, $A \subset \mathbb{R}^{k_1}$, $K \subset \mathbb{R}^{k_2}$, $k_1, k_2 \in \mathbb{N}$, und sei K kompakt. Dann gelten für alle Teilmengen $B \subset K$:

- (i) $\lim_{x \rightarrow x_0} \sup_{y \in B} |f(x, y) - f(x_0, y)| = 0 \quad \forall x_0 \in A,$
- (ii) $\sup_{y \in B} f(\cdot, y)$ und $\inf_{y \in B} f(\cdot, y)$ sind stetige Funktionen auf A .

Beweis:

(i) Seien $x_0 \in A$ und $\varepsilon > 0$ fest. Wegen der Stetigkeit von $f(\cdot)$ existieren für alle Punkte $y \in K$ eine δ' -Umgebung $S_{\delta'}(x_0)$ um x_0 in A und eine δ' -Umgebung $S_{\delta'}(y)$ um y in K mit

$$\sup_{x \in S_{\delta'}(x_0)} \sup_{u \in S_{\delta'}(y)} |f(x, u) - f(x_0, u)| < \varepsilon.$$

Da K kompakt ist, überdecken endlich viele der $S_{\delta'}(y)$ -Umgebungen die Menge K und daher B - etwa $S_{\delta_1}(y_1), \dots, S_{\delta_k}(y_k)$.

Daher folgt mit $\delta := \min\{\delta_1, \dots, \delta_k\}$

$$\begin{aligned} & \sup_{x \in S_{\delta}(x_0)} \sup_{y \in B} |f(x, y) - f(x_0, y)| \\ & \leq \sup_{x \in S_{\delta}(x_0)} \max_{j \in \{1, \dots, k\}} \sup_{y \in S_{\delta_j}(y_j)} |f(x, y) - f(x_0, y)| \\ & < \varepsilon, \end{aligned}$$

womit (i) bewiesen ist.

(ii) Die erste Behauptung folgt aus (i), da für alle $x, x' \in A$

$$|\sup_{y \in B} f(x, y) - \sup_{y \in B} f(x', y)| = \sup_{y \in B} |f(x, y) - f(x', y)|$$

gilt. Die zweite Behauptung folgt aus der ersten, indem $f(\cdot)$ durch $-f(\cdot)$ ersetzt wird.

Den Beweis von (11.6) führen wir nun in drei Schritten durch (Lemmata 11.3, 11.4 und 11.5).

Lemma 11.3

Sei $f(\cdot)$ eine stetige und beschränkte Funktion $X \rightarrow \mathbb{R}$. Dann gilt

$$\frac{1}{n} \sum_{l=1}^n f(X_l) \xrightarrow{\mathbb{P}\text{-f.s.}} \int f dG.$$

Beweis:

Die Behauptung ist äquivalent zu

$$\int f d\hat{G}_n \xrightarrow{\mathbb{P}\text{-f.s.}} \int f dG.$$

Diese Beziehung folgt aber sofort aus (11.2) und der Tatsache, daß für alle $\omega \in \Omega$ mit

$$\sup_{x \in \mathbb{R}^t} \|\hat{G}_n(x; \omega) - G(x)\| \rightarrow 0$$

nach dem Satz von Helly - Bray (Loève (77), S.184)

$$\int f(x) \hat{G}_n(dx; \omega) \rightarrow \int f(x) G(dx)$$

gilt. □

Lemma 11.4

Sei $f(\cdot)$ eine stetige und beschränkte Funktion $\mathcal{Y} \rightarrow \mathbb{R}$. Dann gilt

$$\frac{1}{n} \sum_{l=1}^n f(Y_l) \xrightarrow{\mathbb{P}\text{-f.s.}} \mathbb{E} \cdot f(Y)_f.$$

Beweis:

Wir betrachten zunächst die Zufallsfolge $f(Y_n)_f$ bedingt auf

$\{X_n\}$. Aufgrund der Beschränktheit von $f(\cdot)$ gilt

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \mathbb{E}\{f^2(Y_n) | X_n\} < \infty \quad \text{P-f.s.}$$

Wegen $\mathbb{P}\{Y_n \in A | X_1, X_2, \dots\} = \mathbb{P}\{Y_n \in A | X_n\}$ P-f.s. $\forall A \in \mathcal{B}^S$ und wegen der bedingten Unabhängigkeit der Zufallsfolge $\{Y_n\}$ gegeben $\{X_n\}$ folgt daher nach dem starken Gesetz der großen Zahlen (Loève (77), S. 253)

$$\mathbb{P}\left\{\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{l=1}^n [f(Y_l) - \mathbb{E}\{f(Y_l) | X_l\}] = 0 | X_1, X_2, \dots\right\} = 1$$

und somit

$$(11.7) \quad \frac{1}{n} \sum_{l=1}^n [f(Y_l) - \mathbb{E}\{f(Y_l) | X_l\}] \xrightarrow{\text{P-f.s.}} 0.$$

Falls $\mathbb{E}\{f(Y_1); x\}$, definiert durch (11.4), eine stetige Funktion von x ist, so gilt außerdem wegen $\mathbb{E}\{f(Y_n) | X_n\} = \mathbb{E}\{f(Y_1); X_n\}$ P-f.s., Lemma 11.3 und (11.5)

$$\frac{1}{n} \sum_{l=1}^n \mathbb{E}\{f(Y_l) | X_l\} \xrightarrow{\text{P-f.s.}} \int_{\mathcal{X}} \mathbb{E}\{f(Y_1); x\} G(dx) = \mathbb{E}\{f(Y)\}.$$

Diese Beziehung und (11.7) liefern dann die Behauptung.

Zu zeigen bleibt die Stetigkeit von $\mathbb{E}\{f(Y_1); x\}$ in x oder äquivalent

$$(11.8) \quad \lim_{x \rightarrow x_0} \int_{\omega} f(w, x) H(dw; x) = \int_{\omega} f(w; x_0) H(dw; x_0) \quad \forall x_0 \in \mathcal{X}.$$

Sei $|f(\cdot)|$ beschränkt durch $S > 0$. Sei ferner $\{x_n\}$ eine Folge aus $\mathcal{X}$ mit $x_n \rightarrow x_0 \in \mathcal{X}$ und sei $\varepsilon > 0$. Dann existiert eine kompakte Menge $K \subset \mathbb{R}^{S-t}$ mit

$$\sup_{n \in N_0} \int_K H(dw; x_n) > 1 - \frac{\varepsilon}{2S}.$$

(Die Klasse der $H(\cdot; x_n)$, $n \in N_0$, zugeordneten Wahrscheinlichkeitsmaße ist nämlich wegen $H(w; x_n) \rightarrow H(w; x_0) \forall w \in \mathbb{R}^{S-t}$ relativ kompakt und daher straff (Billingsley (68), S. 16 und S. 37).)

Somit gilt mit den Abkürzungen $g_n(w) := f(w, x_n)$ und

$$H_n(w) := H(w; x_n) :$$

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \left| \int_{\omega} g_n dH_n - \int_{\omega} g_0 dH_0 \right| \\ & \leq \limsup_{n \rightarrow \infty} \int_{\omega} |g_n - g_0| dH_n + \limsup_{n \rightarrow \infty} \left| \int_{\omega} g_0 dH_n - \int_{\omega} g_0 dH_0 \right| \\ & \leq \limsup_{n \rightarrow \infty} \int_{\omega} |g_n - g_0| dH_n \\ & \leq \limsup_{n \rightarrow \infty} \int_{K \cap \omega} |g_n - g_0| dH_n + \limsup_{n \rightarrow \infty} \int_{\omega - K} 2S dH_n \\ & \leq \limsup_{n \rightarrow \infty} \sup_{w \in K \cap \omega} |g_n(w) - g_0(w)| + \varepsilon, \end{aligned}$$

wobei das zweite Ungleichheitszeichen aufgrund des Satzes von Helly - Bray gilt. Da wegen der Stetigkeit von $f(\cdot)$, der Kompaktheit von $K \cap \omega$ (ω ist nach Modellvoraussetzung abgeschlossen) nach Lemma 11.2(i)

$$\lim_{n \rightarrow \infty} \sup_{w \in K \cap \omega} |f(w, x_n) - f(w, x_0)| = 0$$

gilt und $\varepsilon > 0$ beliebig wählbar ist, ist damit (11.8) bewiesen. □

Lemma 11.5

Sei $f(\cdot)$ eine stetige Funktion $Y \rightarrow \mathbb{R}$ und existiere ein $\delta > 0$ derart, daß

$$(11.9) \quad \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{l=1}^n \mathbb{E}(|f(Y_l)|^{1+\delta}) < \infty$$

gilt, so folgen

$$\mathbb{E}(|f(Y)|) < \infty,$$

$$\frac{1}{n} \sum_{l=1}^n f(Y_l) \xrightarrow{\mathbb{P}\text{-f.s.}} \mathbb{E}\{f(Y)\},$$

und

$$\frac{1}{n} \sum_{l=1}^n \mathbb{E}\{f(Y_l)\} \longrightarrow \mathbb{E}\{f(Y)\}.$$

Beweis :

Wegen (11.9) und Lemma 11.1 existiert ein $0 < \delta < 1$ mit

$$\sum_{n=1}^{\infty} \frac{1}{n^{1+\delta}} \mathbb{E}\{|f(Y_n)|^{1+\delta}\} < \infty.$$

Da außerdem $Y_1, Y_2, \dots$ stochastisch unabhängig sind, existiert nach dem starken Gesetz der großen Zahlen (Loève (77), S. 253) ein $\varepsilon > 0$ (nämlich $\varepsilon := \sqrt{1+\delta} - 1$), so daß

$$\frac{1}{n} \sum_{l=1}^n [|f(Y_l)|^{1+\varepsilon} - \mathbb{E}\{|f(Y_1)|^{1+\varepsilon}\}] \xrightarrow{\mathbb{P}\text{-f.s.}} 0$$

gilt. Wegen (11.9) existiert damit eine Schranke $K > 0$, so daß

$$(11.10) \quad \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{l=1}^n |f(Y_l)|^{1+\varepsilon} < K \quad \mathbb{P}\text{-f.s.}$$

oder äquivalent

$$\limsup_{n \rightarrow \infty} \int |f|^{1+\varepsilon} d\hat{Q}_n < K \quad \mathbb{P}\text{-f.s.}$$

gilt, wobei $\hat{Q}_n(\cdot)$ die empirische Verteilungsfunktion von $Y_1, \dots, Y_n$ ist. Hieraus folgt wegen

$$\int_{\{|f| > \alpha\}} |f| d\hat{Q}_n \leq \frac{1}{\alpha^\varepsilon} \int |f|^{1+\varepsilon} d\hat{Q}_n \quad \forall \alpha > 0$$

$$(11.11) \quad \limsup_{n \rightarrow \infty} \int_{\{|f| > \alpha\}} |f| d\hat{Q}_n < \frac{K}{\alpha^\varepsilon} \quad \mathbb{P}\text{-f.s.}$$

(,d.h. $|f(\cdot)|$ ist für $\mathbb{P}$ -fast alle $\omega \in \Omega$ gleichmäßig integrierbar in $\hat{Q}_n(\cdot; \omega)$). Seien nun $f_\alpha(\cdot) := \min\{\alpha, |f(\cdot)|\}$ und $f^\alpha(\cdot) := |f(\cdot)| - f_\alpha(\cdot)$. Dann gibt es wegen (11.11) zu jedem $\gamma > 0$ ein $\alpha = \alpha(\varepsilon, \gamma) > 0$, so daß

$$(11.12) \quad \gamma > \limsup_{n \rightarrow \infty} \int_{\mathcal{Y}} f^{\alpha} d\hat{Q}_n \geq \limsup_{n \rightarrow \infty} \int_{\mathcal{Y}} f^{\alpha \langle c \rangle} d\hat{Q}_n \quad \mathbb{P}\text{-f.s.}$$

gilt, wobei $c > 0$ beliebig ist und $f^{\alpha \langle c \rangle}(\cdot) := \min\{c, f^\alpha(\cdot)\}$ ist. Wegen der Stetigkeit von $f^{\alpha \langle c \rangle}(\cdot)$ und Lemma 11.4 folgt aus (11.12)

$$\gamma \geq \int_Y f^{\alpha \langle c \rangle} dQ \quad \forall c > 0,$$

wobei $Q(\cdot)$ die Verteilungsfunktion von Y ist. Wegen
 $\lim_{c \rightarrow \infty} f^{\alpha \langle c \rangle}(y) \rightarrow f^\alpha(y) \quad \forall y \in Y$ erhalten wir damit nach dem Satz
 von der monotonen Konvergenz (Loève (77), S.125)

$$(11.13) \quad \gamma \geq \int_Y f^\alpha dQ$$

und wegen $\int_Y f_\alpha dQ \leq \alpha$ daher

$$\int_Y |f| dQ < \infty.$$

Damit ist die erste Behauptung gezeigt.

Weiterhin gilt für dieses α

$$(11.14) \quad \left| \int_Y f d\hat{Q}_n - \int_Y f dQ \right| \leq \int_Y f^+ d\hat{Q}_n + \int_Y f^- dQ + \\ + \left| \int_Y g d\hat{Q}_n - \int_Y g dQ \right|,$$

wobei $g(\cdot) := f_\alpha(\cdot) \cdot \text{sign } f(\cdot)$ ist. Da wegen Lemma 11.4

$$\left| \int_Y g d\hat{Q}_n - \int_Y g dQ \right| \xrightarrow{\mathbb{P}\text{-f.s.}} 0$$

gilt, folgt aus (11.12), (11.13) und (11.14)

$$\limsup_{n \rightarrow \infty} \left| \int_Y f d\hat{Q}_n - \int_Y f dQ \right| \leq 2\gamma \quad \mathbb{P}\text{-f.s..}$$

Da $\gamma > 0$ beliebig wählbar ist, ergibt sich hieraus die zweite Behauptung.

Um die dritte Behauptung zu zeigen, definieren wir

$$Z_n = \frac{1}{n} \sum_{l=1}^n f(Y_l), \quad \forall n \in \mathbb{N}. \text{ Dann gilt nach dem eben Bewiesenen}$$

$$Z_n \xrightarrow{\mathbb{P}} \mathbb{E}\{f(Y)\}.$$

Außerdem folgt aus (11.10) und der Hölder - Ungleichung
 (Rao (73), S.55)

$$\limsup_{n \rightarrow \infty} \left[\frac{1}{n} \sum_{l=1}^n |f(Y_l)| \right]^{1+\epsilon} \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{l=1}^n |f(Y_l)|^{1+\epsilon} < K$$

P-f.s..

Damit sind die Z_n gleichmäßig integrierbar und es gilt (Billingsley (68), S.32)

$$\mathbb{E}\{Z_n\} \longrightarrow \mathbb{E}\{f(Y)\},$$

womit die dritte Behauptung bewiesen ist.

Man beachte, daß Lemma 11.3 und Lemma 11.4 und deren Beweis gültig bleiben, wenn wir anstelle der Unabhängigkeit der Y_n nur fordern:

- $Y_1, Y_2, \dots$ sind bedingt auf X_n stochastisch unabhängig,
- $\mathbb{P}\{Y_n \leq y | X_1, X_2, \dots\} = \mathbb{P}\{Y_n \leq y | X_n\}$ P-f.s. $\forall y \in \mathbb{R}, n \in \mathbb{N}$.

Lemma 11.5 bleibt für diesen Fall gültig, wenn zusätzlich (11.9) durch die schärfere Bedingung (11.10) ersetzt wird.

Folgende Regularitätsannahmen setzen wir für die Konsistenz im Allgemeinen Modell voraus:

K1: $\mathcal{G}$ ist kompakt.

Diese Annahme ist nicht ganz befriedigend. Wir diskutieren daher weiter unten eine alternative Annahme.

K2: $\phi(\cdot)$ ist stetig.

Aus K1 und K2 folgt, daß das Infimum in (11.5) in jedem Fall angenommen wird.

K3: Es existiert ein $\epsilon > 0$, so daß

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{l=1}^n \mathbb{E}[\phi^{1+\epsilon}(Y_l, \theta_0)] < \infty$$

gilt.

K4: Für alle $\theta \in \mathcal{G}$ mit $\theta \neq \theta_0$ gilt

$$\mathbb{E}\{\phi(Y, \theta) - \phi(Y, \theta_0)\} > 0.$$

Der Erwartungswert in K_4 ist definiert, falls K_2 und K_3 gelten; dies folgt, da K_2 und K_3 wegen Lemma 11.5 $E\{\phi(Y, \theta_0)\} < \infty$ implizieren und $\phi(\cdot)$ nichtnegativ ist.

Entsprechend der oberen Bemerkung führen wir den Konsistenzbeweis nun ähnlich wie beim herkömmlichen Beweis im unabhängigen, identisch verteilten Fall durch.

Lemma 11.6

Gelten die Regularitätsannahmen $K_1 - K_4$, so existiert für alle $\theta^* \in \Theta$ mit $\theta^* \neq \theta_0$ eine δ -Umgebung

$$(11.15) \quad S_\delta(\theta^*) := \{\theta \in \Theta: \|\theta - \theta^*\| < \delta\},$$

so daß

$$E\left\{ \inf_{\theta \in S_\delta(\theta^*)} \phi(Y, \theta) - \phi(Y, \theta_0) \right\} > 0$$

gilt.

Beweis:

Zunächst gilt wegen K_2 für alle $y \in \mathcal{Y}$, $\theta^* \in \Theta$

$$\inf_{\theta \in S_\delta(\theta^*)} \phi(y, \theta) \uparrow \phi(y, \theta^*) \quad (\delta \downarrow 0).$$

Da außerdem $\phi(\cdot)$ nur nichtnegative Werte annimmt, folgt mit dem Satz von der monotonen Konvergenz

$$(11.16) \quad E\left\{ \inf_{\theta \in S_\delta(\theta^*)} \phi(Y, \theta) \right\} \longrightarrow E\{\phi(Y, \theta^*)\} \quad (\delta \downarrow 0).$$

Da wegen K_2 , K_3 und Lemma 11.5 $E\{\phi(Y, \theta_0)\} < \infty$ gilt, folgt aus (11.16) zusammen mit der Tatsache, daß für je zwei nichtnegative Zufallsvariablen U und V mit $E\{V\} < \infty$ die Gleichung $E\{U - V\} = E\{U\} - E\{V\}$ gilt,

$$E\left\{ \inf_{\theta \in S_\delta(\theta^*)} \phi(Y, \theta) - \phi(Y, \theta_0) \right\} \longrightarrow E\{\phi(Y, \theta^*) - \phi(Y, \theta_0)\} \quad (\delta \downarrow 0).$$

Da im Fall $\theta^* \neq \theta_0$ wegen K_4 $E\{\phi(Y, \theta^*) - \phi(Y, \theta_0)\} > 0$ gilt, erhalten wir somit die Behauptung. □

Lemma 11.7

Gelten die Regularitätsannahmen K1-K4. Dann gibt es für alle $\theta^* \in \Theta$ mit $\theta^* \neq \theta_0$ ein $\delta > 0$, so daß

$$P\left\{\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{l=1}^n \inf_{\theta \in S_\delta(\theta^*)} [\phi(Y_l, \theta) - \phi(Y_l, \theta_0)] > 0\right\} = 1$$

gilt, wobei $S_\delta(\theta^*)$ durch (11.15) definiert ist.

Beweis:

Definiere für irgendeine eindimensionale Zufallsvariable Z $Z^{(c)}$, $c \in \mathbb{R}$, durch $Z^{(c)} := \min\{c, Z\}$. Definiere weiter die Zufallsvariablen

$$(11.17) \quad R_\delta(\theta^*) := \inf_{\theta \in S_\delta(\theta^*)} [\phi(Y, \theta) - \phi(Y, \theta_0)]$$

und

$$(11.18) \quad R_{\delta n}(\theta^*) := \inf_{\theta \in S_\delta(\theta^*)} [\phi(Y_n, \theta) - \phi(Y_n, \theta_0)].$$

Nach Lemma 11.6 existiert für jedes $\theta^* \in \Theta$ mit $\theta^* \neq \theta_0$ ein $\delta > 0$, so daß

$$(11.19) \quad E\{R_\delta(\theta^*)\} > 0$$

gilt. Da $R_\delta^{(c)}(\theta^*) \uparrow R_\delta(\theta^*)$ für $c \uparrow \infty$ und wegen (11.19) $E\{R_\delta^{(c)}(\theta^*)\} > -\infty$ gelten, folgt mit dem Satz von der monotonen Konvergenz

$$\lim_{c \uparrow \infty} E\{R_\delta^{(c)}(\theta^*)\} = E\{R_\delta(\theta^*)\}.$$

Damit existiert ein $c = c(\theta^*, \delta) > 0$, so daß

$$(11.20) \quad E\{R_\delta^{(c)}(\theta^*)\} > 0$$

gilt. Weiter gibt es wegen $c \geq R_{\delta n}^{(c)}(\theta^*) \geq -\phi(Y_n, \theta_0) \quad \forall n \in \mathbb{N}$ und wegen K3 ein $\gamma > 0$ mit

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{l=1}^n E\{|R_{\delta n}^{(c)}(\theta^*)|^{1+\gamma}\} < \infty.$$

Da außerdem $R_{\delta n}^{(c)}(\theta^*)$ als Funktion von Y_n wegen der Stetigkeit von $\phi(\cdot)$ und Lemma 11.2(ii) stetig ist, erhalten wir mit Lemma 11.5

$$(11.21) \quad \frac{1}{n} \sum_{l=1}^n R_{\delta n}^{(c)}(\theta^*) \xrightarrow{\text{P-f.s.}} \mathbb{E}\{R_{\delta}^{(c)}(\theta^*)\}.$$

Die Behauptung folgt nun aus $R_{\delta n}(\theta^*) \geq R_{\delta n}^{(c)}(\theta^*)$, (11.20) und (11.21).

□

Satz 11.8

Sind die Regularitätsbedingungen K1-K4 erfüllt, so ist der Schätzer $\hat{\theta}_n$ bzgl. $\phi(\cdot)$ (stark) konsistent für θ_0 :

$$\hat{\theta}_n \xrightarrow{\text{P-f.s.}} \theta_0$$

(im Sinne der Definition im Anhang B).

Beweis:

Sei $\varepsilon > 0$ beliebig und seien $S_{\delta}(\theta)$ und $R_{\delta n}(\theta)$ durch (11.15) und (11.18) definiert. Wegen Lemma 11.7 gibt es zu jedem $\theta^* \in \Theta - S_{\varepsilon}(\theta_0)$ eine Umgebung $S_{\delta}(\theta^*)$, so daß

$$(11.22) \quad \mathbb{P}\{\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{l=1}^n R_{\delta n}(\theta^*) > 0\} = 1$$

gilt. Da die Menge $\Theta - S_{\varepsilon}(\theta_0)$ wegen K1 kompakt ist, wird sie von endlich vielen solchen Umgebungen - etwa $S_{\delta_1}(\theta_1), \dots, S_{\delta_m}(\theta_m)$ - überdeckt. Mit (11.22) folgt daher

$$\begin{aligned} & \mathbb{P}\{\liminf_{n \rightarrow \infty} \inf_{\theta \in \Theta - S_{\varepsilon}(\theta_0)} \frac{1}{n} \sum_{l=1}^n [\phi(Y_l, \theta) - \phi(Y_l, \theta_0)] > 0\} \\ & \geq \mathbb{P}\{\liminf_{n \rightarrow \infty} \min_{j \in \{1, \dots, m\}} \frac{1}{n} \sum_{l=1}^n R_{\delta_j l}(\theta_j) > 0\} = 1 \end{aligned}$$

und damit

$$\begin{aligned}
 (11.23) \quad & \lim_{m \rightarrow \infty} \mathbb{P} \left[\bigcap_{n=m}^{\infty} \left\{ \inf_{\theta \in \Theta} \sum_{l=1}^n \phi(Y_l, \theta) < \inf_{\theta \in S_{\varepsilon}(\theta_0)} \sum_{l=1}^n \phi(Y_l, \theta) \right\} \right] \\
 &= \mathbb{P} \left[\bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} \left\{ \inf_{\theta \in \Theta - S_{\varepsilon}(\theta_0)} \sum_{l=1}^n [\phi(Y_l, \theta) - \phi(Y_l, \theta_0)] > 0 \right\} \right] = 1.
 \end{aligned}$$

Da $\phi(y, \theta)$ wegen K2 stetig in θ ist und das Infimum einer stetigen Funktion über einer kompakten Menge angenommen wird, gilt

$$\left\{ \inf_{\theta \in \Theta} \sum_{l=1}^n \phi(Y_l, \theta) < \inf_{\theta \in \Theta - S_{\varepsilon}(\theta_0)} \sum_{l=1}^n \phi(Y_l, \theta) \right\} \subset \{ \hat{\theta}_n \in S_{\varepsilon}(\theta_0) \}.$$

Zusammen mit (11.23) erhalten wir daher

$$\lim_{m \rightarrow \infty} \mathbb{P}_* \left[\bigcap_{n=m}^{\infty} \{ \hat{\theta}_n \in S_{\varepsilon}(\theta_0) \} \right] = 1.$$

Da $\varepsilon > 0$ beliebig ist, ist damit die Behauptung bewiesen. □

Wie im ersten Kapitel ist auch hier $\hat{\theta}_n$ im allgemeinen nicht meßbar. Für die nachfolgenden Konvergenz- und Verteilungsaussagen sind daher wieder die im Anhang B aufgeführten allgemeinen Definitionen zu benutzen.

Wir betrachten nun den oben angeführten Fall, daß Θ nicht kompakt ist. In diesem Fall ist eine zusätzliche Bedingung an $\phi(\cdot)$ erforderlich. In der Literatur wird K1 meist durch (im Prinzip) folgende beiden Bedingungen ersetzt:

- (i) Θ ist abgeschlossen,
- (ii) $\liminf_{\|\theta\| \rightarrow \infty} \phi(y, \theta) > \phi(y, \theta_0) \quad \forall y \in J.$

Die Bedingung (ii) ist für das in dieser Arbeit zugrundegelegte Modell (und für viele andere wichtige Modelle ebenfalls) im allgemeinen nicht erfüllt, so daß wir hierauf nicht weiter eingehen werden. Wir nehmen stattdessen anstelle K1 die Regularitätsbedingungen:

K1a: Θ ist konvex und lokalkompakt.

K1b: $\phi(y, \cdot)$ ist für alle $y \in J$ konvex.

Satz 11.9

Gelten die Regularitätsannahmen K1a, K1b und K2 - K4. Dann folgen

$$\lim_{m \rightarrow \infty} \mathbb{P}_* \left[\bigcap_{n=m}^{\infty} \left\{ \sum_{l=1}^n \phi(Y_l, \hat{\theta}_n) = \inf_{\theta \in \Theta} \sum_{l=1}^n \phi(Y_l, \theta) \right\} \right] = 1$$

und

$$\hat{\theta}_n \xrightarrow{\text{P-f.s.}} \theta_o.$$

Beweis :

Wegen K1a gibt es ein $\varepsilon > 0$, so daß $\bar{S}_\varepsilon(\theta_o) := \{\theta \in \Theta : \|\theta - \theta_o\| \leq \varepsilon\}$ kompakt ist. Sieht man $\bar{S}_\varepsilon(\theta_o)$ als neuen Parameterraum Θ an, so sind die Bedingungen für Satz 11.8 erfüllt und es gilt wegen (11.23)

$$(11.24) \quad \lim_{m \rightarrow \infty} \mathbb{P} \left[\bigcap_{n=m}^{\infty} \left\{ \inf_{\theta \in \bar{S}_\varepsilon(\theta_o) - S_{\varepsilon/2}(\theta_o)} \sum_{l=1}^n [\phi(Y_l, \theta) - \phi(Y_l, \theta_o)] > 0 \right\} \right] = 1.$$

Wir zeigen als nächstes, daß aus

$$\theta \in \bar{S}_\varepsilon(\theta_o) - S_{\varepsilon/2}(\theta_o) \quad \sum_{l=1}^n \phi(Y_l, \theta) > \sum_{l=1}^n \phi(Y_l, \theta_o)$$

und aus K1a, K1b die Ungleichung

$$(11.25) \quad \inf_{\theta \in \Theta - S_\varepsilon(\theta_o)} \sum_{l=1}^n \phi(Y_l, \theta) \geq \inf_{\theta \in \bar{S}_\varepsilon(\theta_o) - S_{\varepsilon/2}(\theta_o)} \sum_{l=1}^n \phi(Y_l, \theta)$$

folgt. Sei dazu für festes $\omega \in \Omega$ und $n \in \mathbb{N}$ $f(\theta) := \sum_{l=1}^n \phi(Y_l(\omega), \theta)$.

Sei ferner $\tilde{\theta} \in \Theta - S_\varepsilon(\theta_o)$. Dann folgt aus der Konvexität von $f(\cdot)$ und $\tilde{\theta}$

$$\alpha f(\tilde{\theta}) + (1-\alpha) f(\theta_o) = f(\alpha \tilde{\theta} + (1-\alpha) \theta_o) \quad \forall \alpha \in [0, 1].$$

Wählen wir nun $\alpha > 0$ derart, daß $\bar{\theta} := \alpha \tilde{\theta} + (1-\alpha) \theta_o \in \bar{S}_\varepsilon(\theta_o) - S_{\varepsilon/2}(\theta_o)$ gilt, so erhalten wir wegen $f(\bar{\theta}) > f(\theta_o)$

$$\alpha f(\tilde{\theta}) + (1-\alpha) f(\bar{\theta}) \geq f(\bar{\theta})$$

oder

$$f(\tilde{\theta}) \geq f(\bar{\theta}) \geq \inf_{\theta \in \bar{S}_\varepsilon(\theta_0) - S_{\varepsilon/2}(\theta_0)} f(\theta)$$

und (11.25) folgt, da $\hat{\theta} \in \Theta - S_\varepsilon(\theta_0)$ beliebig ist. Somit erhalten wir aus (11.24) und (11.25)

$$\lim_{m \rightarrow \infty} \mathbb{P} \left[\bigcap_{n=m}^{\infty} \left\{ \inf_{\theta \in \Theta - S_\varepsilon(\theta_0)} \sum_{l=1}^n [\phi(Y_l, \theta) - \phi(Y_l, \theta_0)] > 0 \right\} \right] = 1.$$

Wegen

$$\begin{aligned} & \left\{ \inf_{\theta \in \Theta - S_\varepsilon(\theta_0)} \sum_{l=1}^n [\phi(Y_l, \theta) - \phi(Y_l, \theta_0)] > 0 \right\} \\ & \subset \left\{ \inf_{\theta \in S_\varepsilon(\theta_0)} \sum_{l=1}^n \phi(Y_l, \theta) = \sum_{l=1}^n \phi(Y_l, \hat{\theta}_n) < \right. \\ & \quad \left. < \inf_{\theta \in \Theta - S_\varepsilon(\theta_0)} \sum_{l=1}^n \phi(Y_l, \theta) \right\} \\ & \subset \{ \hat{\theta}_n \in S_\varepsilon(\theta_0) \}. \end{aligned}$$

sind damit beide Behauptungen bewiesen. □

Wir kommen nun zu unserem Kontingenztafelmodell zurück. Wir definieren $\mathcal{M} := \{0, \dots, M\}^{2k}$, $W_n := (N_n, D_n)$, $n \in \mathbb{N}$, und

$$\begin{aligned} (11.26) \quad \phi((m, d, x), \theta) &:= \mathcal{D}(p, \rho(x, \theta), m) \\ &= \sum_{i=1}^k m_i h_i(p_i, \rho_i(x, \theta)) \quad \forall m, d \in \{0, \dots, M\}^k, \\ &\quad x \in \mathcal{X}, \theta \in \Theta, \end{aligned}$$

wobei entsprechend (10.1)

$$p_i := \begin{cases} d_i / m_i, & \text{falls } m_i > 0, d_i \leq m_i \\ 0, & \text{sonst} \end{cases}, \quad i=1, \dots, k.$$

$(d_i, m_i, p_i, \rho_i(\cdot))$ bezeichnen die Komponenten von $d, m, p, \rho(\cdot)$ gelten.

Wir setzen die Gültigkeit folgender Regularitätsbedingungen voraus:

K1': Θ ist kompakt.

K2': Die Funktionen $h_i(\cdot)$, $i=1, \dots, k$, sind stetig.

K3': Für alle $\frac{1}{m} \mathcal{P}(m, \pi)$ -verteilten Zufallsvariablen P ($m \in \mathbb{N}$, $\pi \in (0, 1)$ beliebig) gilt:

- (i) $E\{h_i(P, a) - h_i(P, \pi)\} > 0 \quad \forall a \in (0, 1) \text{ mit } a \neq \pi, i=1, \dots, k,$
- (ii) es existiert eine Schranke $K > 0$, unabhängig von π und m , mit $E\{h_i^2(P, \pi)\} < K, i=1, \dots, k.$

K4': Für alle $\theta \in \Theta$ mit $\theta \neq \theta_0$ gilt

$$\int I_{\{x: \phi(x, \theta) \neq \phi(x, \theta_0)\}}(x) G(dx) > 0.$$

Die Annahme K1' können wir wieder ersetzen durch:

K1'a: Θ ist konvex und lokalkompakt.

K1'b: Die Funktionen $h_i(p, \rho_i(x, \cdot))$, $i=1, \dots, k$, sind für alle $p \in [0, 1]$, $x \in \mathcal{X}$ konvex.

Der nachfolgende Satz beweist, daß aus der Gültigkeit dieser Regularitätsannahmen die Gültigkeit der Regularitätsbedingungen für das Allgemeine Modell folgen.

Satz 11.10

Es gelten folgende Implikationen:

$$\begin{aligned} K1' &\implies K1, \\ K1'a, b &\implies K1a, b, \\ K2' &\implies K2, \\ K3'(ii) &\implies K3, \\ K3'(i), K4' &\implies K4. \end{aligned}$$

Beweis:

Die ersten drei Implikationen sind trivial.

Da $h_i(\cdot)$, $i=1, \dots, k$, nur nichtnegative Werte annehmen und die Zufallsvariablen N_{ni} , $i=1, \dots, k$, $n \in \mathbb{N}$, gleichmäßig durch $M > 0$ beschränkt sind, gilt für alle $x \in \mathcal{X}$, $n \in \mathbb{N}$

$$\begin{aligned} \mathbb{E}\{\mathcal{D}^2(P_n, \rho(X_n, \theta_0), N_n); x\} &\leq M^2 \mathbb{E}\left\{\left[\sum_{i=1}^k h_i(P_{ni}, \rho_i(X_n, \theta_0))\right]^2; x\right\} \\ &\leq M^2 k \mathbb{E}\left\{\sum_{i=1}^k h_i^2(P_{ni}, \rho_i(X_n, \theta_0)); x\right\}, \end{aligned}$$

wobei $\mathbb{E}\{\cdot; x\}$ durch (8.4) definiert ist und die zweite Ungleichung aufgrund der Cauchy - Schwarz - Ungleichung (Rao (73), S.54) folgt. Wegen $K3'(ii)$ und Modellannahme II erhalten wir somit die Existenz einer Schranke $S > 0$ mit

$$\begin{aligned} \mathbb{E}\{\mathcal{D}^2(P_n, \rho(X_n, \theta_0), N_n)\} &= \int \mathbb{E}\{\mathcal{D}^2(P_n, \rho(X_n, \theta_0), N_n); X_n\} dP \\ &< S \quad \forall n \in \mathbb{N}. \end{aligned}$$

Damit ist die vierte Implikation gezeigt.

Gelten nun $K3'(i)$ und $K4'$. Sei $\theta \in \Theta$ mit $\theta \neq \theta_0$ beliebig und sei $A := \{x \in X : \rho(x, \theta) \neq \rho(x, \theta_0)\}$. Dann folgt mit $K3'(i)$ und den Modellbedingungen II und IV die Beziehung

$$\begin{aligned} &\mathbb{E}\{N_{1i}[h_i(P_{1i}, \rho_i(X_1, \theta)) - h_i(P_{1i}, \rho_i(X_1, \theta_0))]; x\} \\ &\begin{cases} > 0, & \text{falls } \rho_i(x, \theta) \neq \rho_i(x, \theta_0) \\ = 0, & \text{sonst} \end{cases} \quad \forall x \in X, i=1, \dots, k. \end{aligned}$$

Hieraus erhalten wir

$$\begin{aligned} &\mathbb{E}\{\mathcal{D}(P_1, \rho(X_1, \theta), N_1) - \mathcal{D}(P_1, \rho(X_1, \theta_0), N_1); x\} \\ &\begin{cases} > 0, & \text{falls } x \in A \\ = 0, & \text{sonst} \end{cases} \quad \forall x \in X \end{aligned}$$

und daher mit $K4'$ und (11.5) die letzte Implikation. □

Da in diesem Beweis die Modellannahme III nicht benötigt wird, ist sie für die Konsistenz nicht notwendig. (Die Annahme III entspricht der Modellannahme II aus Kapitel I, welche ebenfalls für die Konsistenz nicht benötigt wurde.)

Wir diskutieren nun die Gültigkeit der Regularitätsannahme $K3'$ für den Maximum - Likelihood - Schätzer, den Minimum - χ^2 - Schätzer und den Einfachen Schätzer.

Maximum - Likelihood - Schätzer :

Die Abstandsfunktionen sind

$$h_i(p, a) := -p \log a - (1-p) \log(1-a), \quad p \in [0, 1], \quad a \in (0, 1), \\ i=1, \dots, k.$$

Für alle $a \in (0, 1)$ und jede $\frac{1}{m} \mathcal{L}(m, \pi)$ -verteilte Zufallsvariable P ($m \in \mathbb{N}$, $0 < \pi < 1$) gilt somit

$$\mathbb{E}\{h_1(P, a)\} = -\pi \log a - (1-\pi) \log(1-a).$$

Hieraus folgt $K3'(i)$ mit der Tatsache, daß die Funktion $f_\pi(a) := -\pi \log a - (1-\pi) \log(1-a)$ an der Stelle $a = \pi$ ein striktes Minimum besitzt. $K3'(ii)$ folgt aus

$$\begin{aligned} \text{Var}\{h_1(P, \pi)\} &= \text{Var}\left\{P \log \frac{1-\pi}{\pi}\right\} \\ &= \frac{\pi(1-\pi)}{m} \left(\log \frac{1-\pi}{\pi}\right)^2 \end{aligned}$$

und der Tatsache, daß $\pi(1-\pi) \left(\log \frac{1-\pi}{\pi}\right)^2 \rightarrow 0$ für $\pi \rightarrow 0$ und für $\pi \rightarrow 1$ gilt.

Minimum - χ^2 - Schätzer :

Die Abstandsfunktionen sind

$$h_i(p, a) := \frac{(p-a)^2}{a(1-a)}, \quad p \in [0, 1], \quad a \in (0, 1), \quad i=1, \dots, k.$$

Annahme $K3'(i)$ ist hier nicht erfüllt, denn für alle $a \in (0, 1)$ und jede $\mathcal{L}(1, \pi)$ -verteilte Zufallsvariable P ($0 < \pi < 1$) gilt

$$\begin{aligned} \mathbb{E}\{h_i(P, a) - h_i(P, \pi)\} &= \mathbb{E}\left\{\frac{(P-a)^2}{a(1-a)} - \frac{(P-\pi)^2}{\pi(1-\pi)}\right\} \\ &= \frac{\pi^2 - 2a\pi + a^2}{a(1-a)} - 1 \\ &= (\pi - a)\left(\frac{1}{a} - \frac{1}{1-a}\right), \end{aligned}$$

wobei benutzt wurde, daß $E\{P\} = E\{P^2\} = \pi$ gilt. $(\pi - a)(\frac{1}{a} - \frac{1}{1-a})$ kann aber nicht für alle $a \in (0,1)$ größer als 0 sein.

Daß der Minimum- χ^2 -Schätzer tatsächlich im allgemeinen nicht konsistent ist, zeigt das folgende einfache Beispiel: Seien $k=1$ und $N_{n1} \equiv 1$, $X_n \equiv 0$ und $\mathcal{L}_{\theta}(D_{n1}) = \mathcal{G}(1, \theta)$ mit $\theta \in \Theta := (0,1)$ für alle $n \in \mathbb{N}$. $\hat{\theta}_n$ ist die Stelle, an der die Funktion

$$f(\theta) := \sum_{l=1}^n \frac{(P_l - \theta)^2}{\theta(1-\theta)}$$

$(P_l = D_{l1})$ ihr Minimum annimmt. Ableitung von $f(\cdot)$ nach θ liefert $\hat{\theta}_n$ als Nullstelle von

$$\frac{\partial}{\partial \theta} f(\theta) = \frac{1}{n} \sum_{l=1}^n \left\{ 2 \frac{\theta - P_l}{\theta(1-\theta)} - \left[\frac{\theta - P_l}{\theta(1-\theta)} \right]^2 (1-2\theta) \right\}.$$

Unter dem wahren Parameter $\theta_0 \in (0,1)$ gilt dann wegen

$$\frac{1}{n} \sum_{l=1}^n P_l \xrightarrow{\mathbb{P}\text{-f.s.}} \theta_0 \quad \text{und} \quad \frac{1}{n} \sum_{l=1}^n P_l^2 \xrightarrow{\mathbb{P}\text{-f.s.}} \theta_0$$
 die Beziehung

$$\hat{\theta}_n \xrightarrow{\mathbb{P}\text{-f.s.}} \begin{cases} \frac{1}{2}, & \text{falls } \theta_0 = \frac{1}{2} \\ \frac{\theta_0}{1-2\theta_0} \left(\sqrt{\frac{1-\theta_0}{\theta_0}} - 1 \right), & \text{falls } \theta_0 \in (0,1) - \left\{ \frac{1}{2} \right\}, \end{cases}$$

so daß $\hat{\theta}_n$ für den Fall $\theta_0 \neq \frac{1}{2}$ nicht konsistent ist.

Einfacher Schätzer:

Die Abstandsfunktionen sind

$$h_i(p, a) := c_i(p-a)^2, \quad p \in [0,1], \quad a \in (0,1), \quad i=1, \dots, k,$$

wobei $c_i > 0$, $i=1, \dots, k$, vorgegebene Konstanten (Gewichte) sind. Ist P eine $\frac{1}{m} \mathcal{B}(m, \pi)$ -verteilte Zufallsvariable, $m \in \mathbb{N}$, $0 < \pi < 1$, so gilt für alle $a \in (0,1)$ mit $a \neq \pi$, $i=1, \dots, k$

$$\begin{aligned}\mathbb{E}\{h_i(P,a) - h_i(P,\pi)\} &= c_i \mathbb{E}\{(P-a)^2 - (P-\pi)^2\} \\ &= c_i (a - \pi)^2 \\ &> 0\end{aligned}$$

und

$$\mathbb{E}\{h_i^2(P,\pi)\} \leq c_i^2$$

und daher K3'.

Die Annahme K1'b wird beim Maximum-Likelihood-Schätzer erfüllt, falls für alle $x \in X$, $i=1, \dots, k$ sowohl $\log \rho_i(x, \cdot)$ als auch $\log[1 - \rho_i(x, \cdot)]$ konkav ist. Dies ist insbesondere der Fall, wenn $\rho_i(x, \cdot)$ die Form

$$\rho_i(x, \theta) := F(\theta_1 x_{i1} + \dots + \theta_r x_{ir}) \text{ mit } x = (x_{11}, \dots, x_{1r}, \dots, x_{kr})'$$

besitzt und $F(\cdot)$ die Verteilungsfunktion der Normalverteilung, der logistischen oder der inversen log-log-Verteilung ist. (Beweis durch Bilden der zweiten Ableitung von $\log F(\cdot)$ und $\log[1 - F(\cdot)]$, wobei im Fall der Normalverteilung die Beziehung $\int_{-\infty}^x \exp(-\frac{u^2}{2}) du < -\frac{1}{x} \exp(-\frac{x^2}{2})$, falls $x < 0$, (siehe Feller (68, S.175)) ausgenutzt wird.)

12. Asymptotische Normalität

Wir behandeln zunächst wieder das Allgemeine Modell und zeigen die asymptotische Normalität des Schätzers bzgl. einer allgemeinen Funktion $\phi(\cdot)$. Hierzu benötigen wir die folgenden Regularitätsannahmen:

N1: Θ ist offen.

N2: Die Funktion $\phi(y, \theta)$ ist für alle $y \in \mathcal{Y}$ zweimal nach θ differenzierbar, und $D_2\phi(\cdot, \cdot)^{+)}$ und $D_2^2\phi(\cdot, \cdot)$ sind stetig.

N3: $E\{D_2\phi(Y_n, \theta_0)'\} = 0 \quad \forall n \in \mathbb{N}$.

N4: Es existiert ein $\delta > 0$, so daß

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{l=1}^n E\{\|D_2\phi(Y_l, \theta_0)\|^{2+\delta}\} < \infty$$

gilt.

N5: Es existieren ein $\varepsilon > 0$ und ein $\delta > 0$, so daß

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{l=1}^n E\left\{ \sup_{\theta \in S_\varepsilon(\theta_0)} \|D_2^2\phi(Y_l, \theta)\|^{1+\delta} \right\} < \infty \quad ++)$$

gilt.

N6: Die $r \times r$ -Matrizen

$$(12.1) \quad B := E\{D_2\phi(Y, \theta_0) D_2\phi(Y, \theta_0)'\}$$

und

$$(12.2) \quad J := E\{D_2^2\phi(Y, \theta_0)\}$$

sind positiv definit.

Um die Kovarianzmatrix der Limesverteilung des Schätzers konsistent schätzen zu können, benötigen wir außerdem die folgende Verschärfung von N4:

+) $D_2\phi(y, \theta)$ ist ein Zeilenvektor.

++) $S_\varepsilon(\theta_0)$ ist durch (11.15) definiert.

N4a: Es existieren ein $\varepsilon > 0$ und ein $\delta > 0$, so daß

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{l=1}^n \mathbb{E} \left\{ \sup_{\theta \in S_\varepsilon(\theta_0)} \|D_2 \phi(Y_l, \theta)\|^{2+\delta} \right\} < \infty$$

gilt.

Für die nachfolgenden beiden Sätze muß $\hat{\theta}_n$ nicht unbedingt ein Schätzer bzgl. $\phi(\cdot)$ sein, sondern es reicht, wenn $\hat{\theta}_n$ irgendeine (nicht notwendig meßbare) Variable $\Omega \rightarrow \Theta$ mit den Eigenschaften

$$(12.3) \quad \hat{\theta}_n \xrightarrow{\mathbb{P}\text{-f.s.}} \theta_0 \quad +)$$

und

$$(12.4) \quad \frac{1}{\sqrt{n}} \sum_{l=1}^n D_2 \phi(Y_l, \hat{\theta}_n)' \xrightarrow{\mathbb{P}\text{-f.s.}} 0$$

ist. Gelten die Voraussetzungen des Satzes 11.8 oder des Satzes 11.9 und N2 und liegt θ_0 im Innern Θ^0 von Θ , so folgt (12.4) wegen

$$\left\{ \sum_{l=1}^n \phi(Y_l, \hat{\theta}_n) = \inf_{\theta \in \Theta} \sum_{l=1}^n \phi(Y_l, \theta) \right\} \cap \{ \hat{\theta}_n \in \Theta^0 \}$$

$$\subset \left\{ \sum_{l=1}^n D_2 \phi(Y_l, \hat{\theta}_n)' = 0 \right\}.$$

(Damit N1 gültig ist, wird man hier natürlich Θ^0 als neuen Parameterraum wählen.)

Satz 12.1

Sind die Regularitätsbedingungen N1 - N6 erfüllt, so gilt

$$(12.5) \quad \frac{1}{\sqrt{n}} \sum_{l=1}^n D_2 \phi(Y_l, \theta_0)' \xrightarrow{\mathcal{L}} \mathcal{N}(0, B),$$

wobei B durch (12.1) definiert ist. Erfüllt die Folge $\{\hat{\theta}_n\}$ die Bedingungen (12.3) und (12.4), so existieren weiter

$r \times r$ -Matrizen $J_n(\omega)$, $\omega \in \Omega$, $n \in \mathbb{N}$, mit $J_n \xrightarrow{\text{P-f.s.}} J$, wobei J durch (12.2) definiert ist, so daß

$$(12.6) \quad J_n \sqrt{n} (\hat{\theta}_n - \theta_0) + \frac{1}{\sqrt{n}} \sum_{l=1}^n D_2 \phi(Y_l, \theta_0)' \xrightarrow{\text{P-f.s.}} 0. \quad +)$$

gilt. Außerdem folgen

$$(12.7) \quad \sqrt{n} (\hat{\theta}_n - \theta_0) + J^{-1} \frac{1}{\sqrt{n}} \sum_{l=1}^n D_2 \phi(Y_l, \theta_0)' \xrightarrow{P} 0$$

und

$$(12.8) \quad \sqrt{n} (\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, J^{-1} B J^{-1}).$$

Beweis:

Sei λ ein Spaltenvektor aus $\mathbb{R}^r$ mit $\lambda \neq 0$. Wegen N2, N4 und Lemma 11.5 existiert ein $\delta > 0$ mit

$$\frac{1}{n} \sum_{l=1}^n \mathbb{E} \{ |D_2 \phi(Y_l, \theta_0) \lambda|^{2+\delta} \} \longrightarrow \mathbb{E} \{ |D_2 \phi(Y, \theta_0) \lambda|^{2+\delta} \} < \infty.$$

Da wegen N6 $\mathbb{E} \{ [D_2 \phi(Y, \theta_0) \lambda]^2 \} > 0$ gilt, ergibt sich daher

$$\frac{\sum_{l=1}^n \mathbb{E} \{ |D_2 \phi(Y_l, \theta_0) \lambda|^{2+\delta} \}}{\left[\sum_{l=1}^n \mathbb{E} \{ [D_2 \phi(Y_l, \theta_0) \lambda]^2 \} \right]^{1+\delta/2}} \longrightarrow 0.$$

Wegen N3 ist damit die Voraussetzung des zentralen Grenz-

+) (12.6) wird im Abschnitt 13 benötigt.

wertsatzes von Ljapunoff (Loève (77), S. 287) erfüllt und es folgt

$$\frac{\frac{1}{\sqrt{n}} \sum_{l=1}^n D_2 \phi(Y_l, \theta_0) \wedge}{\sqrt{\frac{1}{n} \sum_{l=1}^n \mathbb{E}\{[D_2 \phi(Y_l, \theta_0) \wedge]^2\}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Da aus N2, N4 und Lemma 11.5

$$\frac{1}{n} \sum_{l=1}^n \mathbb{E}\{[D_2 \phi(Y_l, \theta_0) \wedge]^2\} \longrightarrow \lambda' B \lambda$$

folgt, erhalten wir somit

$$\frac{1}{\sqrt{n}} \sum_{l=1}^n D_2 \phi(Y_l, \theta_0) \wedge \xrightarrow{\mathcal{L}} \mathcal{N}(0, \lambda' B \lambda).$$

Nach dem Satz von Cramer - Wold (Billingsley (68), S. 49) folgt hieraus die Behauptung (12.5).

Der übrige Teil des Beweises wird genau wie im unabhängigen und identischen Fall durchgeführt. Wir definieren

$$Z_n(\theta) := \frac{1}{n} \sum_{l=1}^n D_2 \phi(Y_l, \theta)',$$

$$T_n := \frac{1}{n} \sum_{l=1}^n D_2^2 \phi(Y_l, \theta_0),$$

$$L_n(\theta) := \frac{1}{n} \sum_{l=1}^n [D_2^2 \phi(Y_l, \theta) - D_2^2 \phi(Y_l, \theta_0)].$$

Taylorentwicklung von $Z_n(\hat{\theta}_n)$ um θ_0 liefert

$$(12.9) \quad Z_n(\hat{\theta}_n)' = Z_n(\theta_0) + (T_n + L_n(\bar{\theta}_n))(\hat{\theta}_n - \theta_0),$$

wobei $\bar{\theta}_n := \theta_0 + \alpha_n(\hat{\theta}_n - \theta_0)$ mit $0 \leq \alpha_n(\omega) \leq 1 \quad \forall \omega \in \Omega$ ist. Mit (12.4) folgt aus (12.9)

$$(12.10) \quad \sqrt{n} Z_n(\theta_0) + (T_n + L_n(\bar{\theta}_n)) \sqrt{n} (\hat{\theta}_n - \theta_0) \xrightarrow{\mathbb{P}\text{-f.s.}} 0.$$

Als nächstes zeigen wir

$$(12.11) \quad T_n \xrightarrow{\text{P-f.s.}} J$$

und

$$(12.12) \quad L_n(\bar{\theta}_n) \xrightarrow{\text{P-f.s.}} 0.$$

(12.11) folgt sofort aus N2, N5 und Lemma 11.5. Um (12.12) zu beweisen, definieren wir

$$B(\varepsilon) := \sup_{\theta \in \bar{S}_\varepsilon(\theta_0)} \|D_2^2 \phi(Y, \theta) - D_2^2 \phi(Y, \theta_0)\|$$

und $B_n(\varepsilon)$ entsprechend mit Y_n anstelle Y . Wegen N2 gilt dann

$$(12.13) \quad \lim_{\varepsilon \rightarrow 0} B(\varepsilon) = 0.$$

Wegen N1 und N5 existieren außerdem ein $\tilde{\varepsilon} > 0$ und ein $\delta > 0$ derart, daß $\bar{S}_{\tilde{\varepsilon}}(\theta_0)$ kompakt ist und $\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{l=1}^n E\{B_l^{1+\delta}(\tilde{\varepsilon})\} < \infty$ gilt, woraus wegen N2 mit Lemma 11.2(ii) und Lemma 11.5

$$(12.14) \quad E\{B(\tilde{\varepsilon})\} < \infty$$

und

$$(12.15) \quad \frac{1}{n} \sum_{l=1}^n B_l(\varepsilon) \xrightarrow{\text{P-f.s.}} E\{B(\varepsilon)\} \quad \forall 0 < \varepsilon \leq \tilde{\varepsilon}$$

folgen. Wegen (12.13) und (12.14) läßt sich der Satz von der majorisierten Konvergenz (Loève (77), S.16) anwenden; er liefert für jedes $\gamma > 0$ die Existenz eines $0 < \varepsilon \leq \tilde{\varepsilon}$ mit

$$(12.16) \quad E\{B(\varepsilon)\} < \gamma.$$

Da mit $\hat{\theta}_n \xrightarrow{\text{P-f.s.}} \theta_0$ auch $\bar{\theta}_n \xrightarrow{\text{P-f.s.}} \theta_0$ gilt, erhalten wir

$$\lim_{m \rightarrow \infty} P_* \left[\bigcap_{n=m}^{\infty} \|\bar{\theta}_n - \theta_0\| < \varepsilon \right] = 1$$

und daher

$$\lim_{m \rightarrow \infty} P_* \left[\bigcap_{n=m}^{\infty} \{\|L_n(\bar{\theta}_n)\| \leq \frac{1}{n} \sum_{l=1}^n B_l(\cdot)\} \right] = 1$$

und mit (12.16)

$$\lim_{m \rightarrow \infty} \mathbb{P}_* \left[\bigcap_{n=m}^{\infty} \{ \|L_n(\hat{\theta}_n)\| < \frac{1}{n} \sum_{l=1}^n B_l(\varepsilon) - \mathbb{E}\{B(\varepsilon)\} + \gamma \} \right] = 1.$$

Da wegen (12.15) $\frac{1}{n} \sum_{l=1}^n B_l(\varepsilon) \xrightarrow{\mathbb{P}\text{-f.s.}} \mathbb{E}\{B(\varepsilon)\}$ gilt und da $\gamma > 0$ beliebig gewählt werden kann, erhalten wir damit (12.12). (12.11) und (12.12) in (12.10) eingesetzt, liefert die Behauptung (12.6) und damit insbesondere

$$(J + o_{\mathbb{P}}(1)) \sqrt{n} (\hat{\theta}_n - \theta_0) + \sqrt{n} Z_n(\theta_0) \xrightarrow{\mathbb{P}} 0.$$

Da eine Determinante als stetige Funktion ihrer Elemente aufgefaßt werden kann und J wegen N6 invertierbar ist, folgt

$$\begin{aligned} \sqrt{n} (\hat{\theta}_n - \theta_0) &= -(J^{-1} + o_{\mathbb{P}}(1)) \sqrt{n} Z_n(\theta_0) + o_{\mathbb{P}}(1) \\ &= -J^{-1} \sqrt{n} Z_n(\theta_0) + o_{\mathbb{P}}(1). \end{aligned}$$

Damit ist (12.7) bewiesen. (12.8) ist eine Folgerung aus (12.5) und (12.7). □

Gilt

$$\mathbb{E}\{D_2\phi(Y, \theta_0) D_2\phi(Y, \theta_0)'\} = \mathbb{E}\{D_2^2\phi(Y, \theta_0)\},$$

so vereinfacht sich (12.8) zu

$$\sqrt{n} (\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, J^{-1}).$$

Diese Bedingung ist meist erfüllt, wenn $-\phi(\cdot)$ eine volle oder partielle log-Likelihoodfunktion ist.

Der nachfolgende Satz gibt Bedingungen an, unter denen sich die Matrizen B und J (stark) konsistent durch

$$\frac{1}{n} \sum_{l=1}^n D_2\phi(Y_l, \theta_n) D_2\phi(Y_l, \theta_n)'$$

und

$$\frac{1}{n} \sum_{l=1}^n D_2^2 \phi(Y_l, \hat{\theta}_n)$$

schätzen lassen.

Satz 12.2

Gelte

$$\hat{\theta}_n \xrightarrow{\mathbb{P}\text{-f.s.}} \theta_0.$$

Dann folgen aus der Gültigkeit von N1, N2 und N5

$$\frac{1}{n} \sum_{l=1}^n D_2^2 \phi(Y_l, \hat{\theta}_n) \xrightarrow{\mathbb{P}\text{-f.s.}} J$$

und aus der Gültigkeit von N1, N2 und N4a

$$\frac{1}{n} \sum_{l=1}^n D_2^2 \phi(Y_l, \hat{\theta}_n) D_2^2 \phi(Y_l, \hat{\theta}_n)' \xrightarrow{\mathbb{P}\text{-f.s.}} B.$$

Beweis :

Zunächst bleibt die Aussage (12.12) richtig, wenn $\bar{\theta}_n$ durch $\hat{\theta}_n$ ersetzt wird. Wir erhalten also

$$L_n(\hat{\theta}_n) = \frac{1}{n} \sum_{l=1}^n [D_2^2 \phi(Y_l, \hat{\theta}_n) - D_2^2 \phi(Y_l, \theta_0)] \xrightarrow{\mathbb{P}\text{-f.s.}} 0.$$

Da wegen N2, N5 und Lemma 11.5

$$\frac{1}{n} \sum_{l=1}^n D_2^2 \phi(Y_l, \theta_0) \xrightarrow{\mathbb{P}\text{-f.s.}} J$$

gilt, ist damit die erste Behauptung bewiesen.

Ähnlich wie (12.12) beweist man mit N4a die Beziehung

$$\frac{1}{n} \sum_{l=1}^n [D_2^2 \phi(Y_l, \hat{\theta}_n) D_2^2 \phi(Y_l, \hat{\theta}_n)' - D_2^2 \phi(Y_l, \theta_0) D_2^2 \phi(Y_l, \theta_0)'] \xrightarrow{\mathbb{P}\text{-f.s.}}$$

Da wegen N2, N4a und Lemma 11.5

$$\frac{1}{n} \sum_{l=1}^n D_2 \phi(Y_l, \theta_0) D_2 \phi(Y_l, \theta_0)' \xrightarrow{\mathbb{P}\text{-f.s.}} B$$

gilt, folgt damit auch die zweite Behauptung. □

Wir kommen nun wieder zu unserem Kontingenztafelmodell zurück, bei dem $Y_n := (N_n, D_n, X_n) \quad \forall n \in \mathbb{N}$ gilt und $\phi(\cdot)$ durch (11.26) definiert ist. Folgende Regularitätsannahmen legen wir zugrunde:

N1': $\textcircled{H}$ ist offen.

N2': Die Ableitungen $D_2 h_i(\cdot, \cdot)$ und $D_2^2 h_i(\cdot, \cdot)$, $i=1, \dots, k$, existieren und sind stetig.

N3': Die Ableitungen $D_2 \rho_i(\cdot, \cdot)$ und $D_2^2 \rho_i(\cdot, \cdot)$, $i=1, \dots, k$, existieren und sind stetig.

N4': Für alle $\frac{1}{m} B(m, \pi)$ -verteilten Zufallsvariablen P ($m \in \mathbb{N}$, $\pi \in (0, 1)$ beliebig) gilt

$$\mathbb{E} \left\{ \frac{\partial}{\partial \pi} h_i(P, \pi) \right\} = 0, \quad i=1, \dots, k.$$

N5': Es existiert ein $\delta > 0$, so daß für $i=1, \dots, k$, $j=1, \dots, r$

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{l=1}^n \mathbb{E} \left\{ \left| \frac{\partial}{\partial \theta_j} h_i(P_{li}, \rho_i(X_l, \theta_0)) \right|^{2+\delta} \right\} < \infty$$

gilt.

N6': Es existieren ein $\varepsilon > 0$ und ein $\delta > 0$, so daß für $i=1, \dots, k$, $j, u=1, \dots, r$

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{l=1}^n \mathbb{E} \left\{ \sup_{\theta \in S_\varepsilon(\theta_0)} \left| \frac{\partial^2}{\partial \theta_j \partial \theta_u} h_i(P_{li}, \rho_i(X_l, \theta)) \right|^{1+\delta} \right\} < \infty$$

gilt.

N7': Die Matrizen mit den Elementen

$$(12.17) \quad \int \mathbb{E} \left\{ \sum_{i=1}^k N_{li}^2 \frac{\partial h_i}{\partial \theta_j}(P_{li}, \rho_i(x, \theta_0)) \frac{\partial h_i}{\partial \theta_u}(P_{li}, \rho_i(x, \theta_0)); x \right\} G(dx),$$

$j, u=1, \dots, r,$

und

$$(12.18) \quad \int \mathbb{E} \left\{ \sum_{i=1}^k N_{1i} \frac{\partial^2 h_i}{\partial \theta_j \partial \theta_u} (P_{1i}, \rho_i(x, \theta_0)); x \right\} G(dx), \quad j, u=1, \dots, r,$$

(wobei $\mathbb{E}\{\cdot; x\}$ durch (8.4) definiert ist) sind positiv definit.

Für die konsistente Schätzung der Kovarianzmatrix benötigen wir außerdem:

N5'a: Es existieren ein $\varepsilon > 0$ und ein $\delta > 0$, so daß für $i=1, \dots, k$, $j=1, \dots, r$

$$\lim_{n \rightarrow \infty} \sup \frac{1}{n} \sum_{l=1}^n \mathbb{E} \left\{ \sup_{\theta \in S_\varepsilon(\theta_0)} \left| \frac{\partial}{\partial \theta_j} h_i(P_{1i}, \rho_i(X_l, \theta)) \right|^{2+\delta} \right\} < \infty$$

gilt.

Aus der Gültigkeit dieser Regularitätsannahmen folgt die Gültigkeit der Regularitätsbedingungen für das Allgemeine Modell, wie der nachfolgende Satz beweist:

Satz 12.3

Es gelten folgende Implikationen:

$$N1' \implies N1,$$

$$N2', N3' \implies N2,$$

$$N4' \implies N3,$$

$$N5' \implies N4,$$

$$N5'a \implies N4a$$

$$N6' \implies N5$$

$$N4', N7' \implies N6, \text{ wobei } B \text{ und } J \text{ gleich den durch (12.17) und (12.18) definierten Matrizen sind.}$$

Beweis:

Die ersten beiden Implikationen sind sofort klar.

N3 folgt aus Modellbedingung II und N4', da für alle $n \in \mathbb{N}$, $j=1, \dots, r$, $i=1, \dots, k$

$$\begin{aligned} & \mathbb{E}\left\{N_{ni} \frac{\partial h_i}{\partial \theta_j}(P_{ni}, \rho_i(X_n, \theta_o))\right\} \\ &= \int \mathbb{E}\left\{N_{ni} \frac{\partial h_i}{\partial \theta_j}(P_{ni}, \rho_i(X_n, \theta_o)); X_n\right\} dP = 0 \end{aligned}$$

gilt.

Weiter gilt nach der Hölder - Ungleichung (Rao (73), S.55) für alle $n \in \mathbb{N}$, $j=1, \dots, r$ und $\delta > 0$

$$\begin{aligned} & \left[\sum_{i=1}^k \left| N_{ni} \frac{\partial h_i}{\partial \theta_j}(P_{ni}, \rho_i(X_n, \theta_o)) \right| \right]^{2+\delta} \\ & \leq k^{2+\delta} \sum_{i=1}^k \left| N_{ni} \frac{\partial h_i}{\partial \theta_j}(P_{ni}, \rho_i(X_n, \theta_o)) \right|^{2+\delta}. \end{aligned}$$

Hieraus folgt wegen der gleichmäßigen Beschränktheit der N_{ni} die Implikation $N5' \implies N4$. $N5'a \implies N4a$ und $N6' \implies N5$ beweist man analog.

Die erste Bedingung von $N6$ verifiziert man, indem man für jedes $x \in \mathcal{X}$ nacheinander die bedingten Erwartungswerte bzgl. P^x von

$$\left[\sum_{i=1}^k N_{1i} \frac{\partial h_i}{\partial \theta_j}(P_{1i}, \rho_i(x, \theta_o)) \right] \cdot \left[\sum_{i=1}^k N_{1i} \frac{\partial h_i}{\partial \theta_u}(P_{1i}, \rho_i(x, \theta_o)) \right],$$

$j, u=1, \dots, r$, gegeben $(N_{11}, D_{11}, \dots, N_{1k}), (N_{11}, D_{11}, \dots, D_{1, k-1}), \dots, (N_{1k})$ bildet, wodurch wegen $N4'$ und den Modellbedingungen II und III die gemischten Summanden verschwinden und B die Form (12.17) erhält. Bei der zweiten Behauptung ist nichts zu zeigen. □

Wir diskutieren nun die Gültigkeit der Regularitätsannahmen $N4' - N7'$ bei dem Maximum - Likelihood - Schätzer und beim Einfachen Schätzer.

Maximum - Likelihood - Schätzer:

Die Abstandsfunktionen $h_i(\cdot)$, $i=1, \dots, k$, sind definiert durch

$$h_i(p, a) := -p \log a - (1-p) \log(1-a), \quad p \in [0, 1], \quad a \in (0, 1).$$

Die Gültigkeit von $N4'$ folgt sofort wegen

$$\frac{\partial h_1}{\partial a}(p, a) = \frac{a - p}{a(1 - a)}.$$

B und J sind identisch und besitzen die Elemente

$$\int_X \mathbb{E} \left\{ \sum_{i=1}^k \frac{N_{ni}}{\rho_i(x, \theta_0) [1 - \rho_i(x, \theta_0)]} \frac{\partial \rho_i}{\partial \theta_j}(x, \theta_0) \frac{\partial \rho_i}{\partial \theta_u}(x, \theta_0); x \right\} G(dx),$$

$j, u = 1, \dots, r.$

Wegen

$$\mathbb{E} \left\{ \frac{N_{ni}}{\rho_i(x, \theta_0) [1 - \rho_i(x, \theta_0)]}; x \right\} > 0 \quad \forall x \in X, i = 1, \dots, k,$$

ist N7' genau dann erfüllt, wenn die Matrix

$$\mathbb{E} \{ D_2 \rho(X, \theta_0)' D_2 \rho(X, \theta_0) \} = \int_X D_2 \rho(x, \theta_0)' D_2 \rho(x, \theta_0) G(dx)$$

positiv definit ist, wobei $\rho(\cdot) := (\rho_1(\cdot), \dots, \rho_k(\cdot))'$ ist. Äquivalent hierzu ist, daß die Zufallsvektoren

$$\frac{\partial \rho_1}{\partial \theta_j}(X, \theta_0), \dots, \frac{\partial \rho_k}{\partial \theta_j}(X, \theta_0)', \quad j = 1, \dots, r,$$

mit positiver Wahrscheinlichkeit linear unabhängig sind. Annahme N5' ist erfüllt, wenn ein $\delta > 0$ existiert, so daß für $i = 1, \dots, k$ und $j = 1, \dots, r$

$$(12.19) \quad \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{l=1}^n \mathbb{E} \left\{ \left| \frac{1}{\rho_i(X_l, \theta_0) [1 - \rho_i(X_l, \theta_0)]} \frac{\partial \rho_i}{\partial \theta_j}(X_l, \theta_0) \right|^{2+\delta} \right\} < \infty$$

gilt. Annahme N6' ist erfüllt, wenn ein $\varepsilon > 0$ und ein $\delta > 0$ existieren, so daß für $i = 1, \dots, k$ und $j, u = 1, \dots, r$

$$(12.20) \quad \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{l=1}^n \mathbb{E} \left\{ \sup_{\theta \in S_\varepsilon(\theta_0)} \left[\left| \frac{\partial^2 \log \rho_i(X_l, \theta)}{\partial \theta_j \partial \theta_u} \right| + \left| \frac{\partial^2 \log [1 - \rho_i(X_l, \theta)]}{\partial \theta_j \partial \theta_u} \right| \right]^{1+\delta} \right\} < \infty$$

gilt. N5'a ist erfüllt, wenn ein $\varepsilon > 0$ und ein $\delta > 0$ existieren, so daß für $i = 1, \dots, k$ und $j = 1, \dots, r$

$$(12.21) \quad \lim_{n \rightarrow \infty} \sup \frac{1}{n} \sum_{l=1}^n \mathbb{E} \left\{ \sup_{\theta \in S_{\varepsilon}(\theta_0)} \left| \frac{1}{\rho_i(x_l, \theta) [1 - \rho_i(x_l, \theta)]} \frac{\partial \rho_i}{\partial \theta_j}(x_l, \theta) \right|^{2+\delta} \right\} < \infty$$

gilt. (12.19), (12.20) und (12.21) sind dann erfüllt, wenn N3' gilt und X kompakt ist. Ansonsten müssen sie für jedes spezielle Modell einzeln überprüft werden.

Beispiel 12.1 (Logitmodell):

Seien $\rho_i(x, \theta) := F(\theta'x)$, $i=1, \dots, k$, mit $\theta = (\theta_1, \dots, \theta_r)' \in \mathbb{R}^r$, $x = (x_1, \dots, x_r)' \in \mathbb{R}^r$ und $F(u) := 1/(1+e^{-u})$, $u \in \mathbb{R}$, (logistische Verteilung). Wegen

$$\frac{\partial F}{\partial \theta_j}(\theta'x) = F(\theta'x) [1 - F(\theta'x)] x_j, \quad j=1, \dots, r,$$

und

$$\begin{aligned} \frac{\partial^2 \log F(\theta'x)}{\partial \theta_j \partial \theta_1} &= - \frac{\partial^2 \log [1 - F(\theta'x)]}{\partial \theta_j \partial \theta_1} \\ &= F(\theta'x) [1 - F(\theta'x)] x_j x_1, \quad j, l=1, \dots, r, \end{aligned}$$

ist eine hinreichende Bedingung für die Gültigkeit von (12.19), (12.20) und (12.21) die Existenz eines $\delta > 0$ mit

$$\lim_{n \rightarrow \infty} \sup \frac{1}{n} \sum_{l=1}^n \mathbb{E} \{ \|X_l\|^{2+\delta} \} < \infty.$$

Beispiel 12.2 (Probitmodell):

Seien $\rho_i(x, \theta) := F(\theta'x)$, $i=1, \dots, k$, mit $\theta \in \mathbb{R}^r$ und $x \in \mathbb{R}^r$ und $F(\cdot)$ die standardisierte Normalverteilung mit Dichte $f(\cdot)$. Dann gilt für alle $u \in \mathbb{R}$

$$\frac{f(u)}{F(u)} < 2$$

und für alle $u < -\sqrt{2}$

$$\frac{f(u)}{F(u)} < \frac{|u|}{1 - 1/u^2} < 2|u|$$

(Feller (68), S. 175). Hieraus folgt

$$(12.22) \quad \left| \frac{1}{F(\theta'x) [1 - F(\theta'x)]} \frac{\partial F}{\partial \theta_j}(\theta'x) \right| < 4 |x_j| \max\{|\theta'x|, 1\}, \quad j=1, \dots, r$$

Außerdem erhalten wir

$$\begin{aligned} \left| \frac{\partial^2 \log F(\theta'x)}{\partial \theta_j \partial \theta_1} \right| &\leq |x_j x_1| \left| \theta'x \frac{f(\theta'x)}{F(\theta'x)} - \left(\frac{f(\theta'x)}{F(\theta'x)} \right)^2 \right| \\ &< 6 |x_j x_1| \max\{1, (\theta'x)^2\}, \quad j, l=1, \dots, r, \end{aligned}$$

und

$$\begin{aligned} \left| \frac{\partial^2 \log [1 - F(\theta'x)]}{\partial \theta_j \partial \theta_1} \right| &= |x_j x_1| \left| \theta'x \frac{f(\theta'x)}{1 - F(\theta'x)} - \left(\frac{f(\theta'x)}{1 - F(\theta'x)} \right)^2 \right| \\ &< 6 |x_j x_1| \max\{1, (\theta'x)^2\}, \quad j, l=1, \dots, r, \end{aligned}$$

und daher

$$\begin{aligned} (12.23) \quad \sup_{\|\theta - \theta_0\| < \varepsilon} \left[\left| \frac{\partial^2 \log F(\theta'x)}{\partial \theta_j \partial \theta_1} \right| + \left| \frac{\partial^2 \log [1 - F(\theta'x)]}{\partial \theta_j \partial \theta_1} \right| \right] \\ < 12 |x_j x_1| \max\{1, (\|\theta_0\| + \varepsilon)^2 \|x\|^2\} \\ < K |x_j x_1| \max\{1, \|x\|^2\} \end{aligned}$$

für eine geeignete Schranke $K > 0$. Wegen (12.22) und (12.23) sind (12.19), (12.20) und (12.21) erfüllt, wenn ein $\delta > 0$ existiert mit

$$\lim_{n \rightarrow \infty} \sup_{\frac{1}{n} \leq \|x\| \leq 1} \mathbb{E} \|X_1\|^{4+\delta} < \infty.$$

Beispiel 12.3 (komplementäres log-log-Modell):

Seien $\rho_i(x, \theta) := F(\theta'x)$, $i=1, \dots, k$, mit $\theta \in \mathbb{R}^r$, $x \in \mathbb{R}^r$ und $F(\cdot)$ definiert durch $F(u) := 1 - \exp(-e^u)$ (komplementäre log-log-Verteilung). Dann existiert wegen $\lim_{u \rightarrow -\infty} e^u / F(u) = 1$ eine Schranke $K_1 > 0$, unabhängig von x und θ , mit

$$\begin{aligned} (12.24) \quad \left| \frac{1}{F(\theta'x) [1 - F(\theta'x)]} \frac{\partial F}{\partial \theta_j}(\theta'x) \right| &= \left| \frac{x_j e^{\theta'x}}{F(\theta'x)} \right| \\ &\leq K_1 |x_j| e^{\theta'x}. \end{aligned}$$

Wegen $\frac{e^u}{[\exp(e^u) - 1]^2} [\exp(e^u) - \exp(e^u + u) - 1] \rightarrow 0 \quad (u \rightarrow -\infty, u \rightarrow \infty)$

existiert außerdem eine Schranke $K_2 > 0$, unabhängig von x und θ , mit

$$\left| \frac{\partial^2 \log F(\theta'x)}{\partial \theta_j \partial \theta_1} \right| = \left| \frac{e^{\theta'x} [\exp(e^{\theta'x}) - \exp(e^{\theta'x} + \theta'x) - 1] x_j x_1}{[\exp(e^{\theta'x}) - 1]^2} \right| \leq K_2 |x_j x_1|, \quad j, l=1, \dots, r.$$

Unter Beachtung von

$$\left| \frac{\partial^2 \log [1 - F(\theta'x)]}{\partial \theta_j \partial \theta_1} \right| = |x_j x_1 e^{\theta'x}|, \quad j, l=1, \dots, r,$$

folgt daher

$$\sup_{\|\theta - \theta_0\| < \varepsilon} \left\{ \left| \frac{\partial^2 \log F(\theta'x)}{\partial \theta_j \partial \theta_1} \right| + \left| \frac{\partial^2 \log [1 - F(\theta'x)]}{\partial \theta_j \partial \theta_1} \right| \right\} \leq (K_1 + K_2) |x_j x_1| \max\{1, \exp[(\|\theta_0\| + \varepsilon)\|x\|]\}, \quad j, l=1, \dots, r.$$

Eine hinreichende Bedingung für (12.20) ist daher die Existenz eines $\delta > 0$, so daß

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{l=1}^n \mathbb{E}^j e^{(\|\theta_0\| + \delta)\|X_1\|} < \infty$$

gilt. Mit (12.24) folgt dann auch die Gültigkeit von (12.19) und (12.21).

Bei allen drei Beispielen ist N_7' dann erfüllt, wenn die Matrix $\mathbb{E}\{XX'\}$ positiv definit ist. Dies folgt aus

$$\frac{\partial F}{\partial \theta_j}(\theta'x) \frac{\partial F}{\partial \theta_1}(\theta'x) = f^2(\theta'x) x_j x_1, \quad j, l=1, \dots, r,$$

wobei $f(\cdot)$ die zu $F(\cdot)$ gehörende Dichte ist.

Einfacher Schätzer :

Die Abstandsfunktionen $h_i(\cdot)$, $i=1, \dots, k$, sind definiert durch

$$h_i(p, a) := c_i (p - a)^2 \quad p \in [0, 1], \quad a \in (0, 1),$$

wobei $c_i > 0$, $i=1, \dots, k$, vorgegebene Konstanten sind. Die Gültigkeit

von N4' folgt sofort wegen

$$\frac{\partial h_i}{\partial a}(p, a) = 2 c_i (a - p), \quad i=1, \dots, k.$$

B und J besitzen die Elemente

$$\int_X \mathbb{E} \left\{ \sum_{i=1}^k 4 c_i^2 N_{1i} \rho_i(X_1, \theta_0) [1 - \rho_i(X_1, \theta_0)] \cdot \right. \\ \left. \cdot \frac{\partial \rho_i}{\partial \theta_j}(X_1, \theta_0) \frac{\partial \rho_i}{\partial \theta_1}(X_1, \theta_0); x \right\} G(dx), \quad j, l=1, \dots, r,$$

und

$$\int_X \mathbb{E} \left\{ \sum_{i=1}^k 2 c_i N_{1i} \frac{\partial \rho_i}{\partial \theta_j}(X_1, \theta_0) \frac{\partial \rho_i}{\partial \theta_1}(X_1, \theta_0); x \right\} G(dx), \quad j, l=1, \dots, r.$$

Wegen $\mathbb{E}\{N_{1i}; x\} > 0 \quad \forall x \in X, i=1, \dots, k$, ist N7' genau dann erfüllt, wenn die Matrix

$$\mathbb{E}\{D_2 \rho(X, \theta_0)' D_2 \rho(X, \theta_0)\}$$

positiv definit ist (,wobei X die Verteilungsfunktion $G(\cdot)$ besitzt.) Annahme N5'a (und daher auch N5') ist erfüllt, wenn ein $\varepsilon > 0$ und ein $\delta > 0$ existieren, so daß für $i=1, \dots, k$ und $j=1, \dots, r$

$$(12.25) \quad \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{l=1}^n \mathbb{E} \left\{ \sup_{\theta \in S_\varepsilon(\theta_0)} \left| \frac{\partial \rho_i}{\partial \theta_j}(X_1, \theta) \right|^{2+\delta} \right\} < \infty$$

gilt, und Annahme N6' ist erfüllt, wenn ein $\varepsilon > 0$ und ein $\delta > 0$ existieren, so daß für $i=1, \dots, k$ und $j, u=1, \dots, r$

$$(12.26) \quad \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{l=1}^n \mathbb{E} \left\{ \sup_{\theta \in S_\varepsilon(\theta_0)} \left[\left| \frac{\partial \rho_i}{\partial \theta_j}(X_1, \theta) \frac{\partial \rho_i}{\partial \theta_u}(X_1, \theta) \right| + \right. \right. \\ \left. \left. + \left| \frac{\partial^2 \rho_i}{\partial \theta_j \partial \theta_u}(X_1, \theta) \right| \right]^{1+\delta} \right\} < \infty.$$

Man rechnet leicht nach, daß für die drei Beispiele eine hinreichende Bedingung für die Gültigkeit von (12.25) und (12.26) die Existenz eines $\delta > 0$ mit

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{l=1}^n \mathbb{E} \{ \|X_n\|^{2+\delta} \} < \infty$$

ist.

13. Konsistenz und asymptotische Normalität bei zufällig verteilter Stichprobenanzahl

Konsistenz und asymptotische Normalität des Schätzers $\hat{\theta}_n$ haben wir unter der Voraussetzung gezeigt, daß $\{v_n\}$ eine feste Folge von Zufallsvektoren ist. Dieser Ansatz ist gerechtfertigt, wenn die Anzahl der Versuchseinheiten vom Experimentator festgelegt wird. Häufig ist diese Anzahl jedoch selbst eine Zufallsvariable, z.B. eine poissonverteilte. Die Frage ist, unter welchen Bedingungen Satz 11.7 und Satz 12.1 gültig bleiben, wenn n durch eine Zufallszahl v_n ersetzt wird. Eine sinnvolle Bedingung für die starke (schwache) Konsistenz von $\{\hat{\theta}_{v_n}\}$ ist sicherlich $v_n \xrightarrow{\text{P-f.s.}} \infty$ ($v_n \xrightarrow{\text{P}} \infty$). Daß diese Bedingungen hinreichend sind, geht sofort aus dem nachfolgenden Lemma hervor, indem man $Z_n = \hat{\theta}_n - \theta_0$ setzt.

Lemma 13.1

Sei $\{Z_n\}$ eine Folge von (nicht notwendig meßbaren) Variablen auf $(\Omega, \mathcal{A}, \mathbb{P})$ und $\{v_n\}$ eine Folge von Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$ mit Werten in $\mathbb{N}$. Dann gelten

$$\begin{aligned} & Z_n \xrightarrow{\text{P-f.s.}} 0, \quad v_n \xrightarrow{\text{P-f.s.}} \infty \implies Z_{v_n} \xrightarrow{\text{P-f.s.}} 0 \\ \text{und} \quad & Z_n \xrightarrow{\text{P-f.s.}} 0, \quad v_n \xrightarrow{\text{P}} \infty \implies Z_{v_n} \xrightarrow{\text{P}} 0. \end{aligned}$$

Beweis :

$Z_n \xrightarrow{\text{P-f.s.}} 0$ ist äquivalent dazu, daß für alle $\eta, \varepsilon > 0$ ein $m \in \mathbb{N}$ mit

$$\mathbb{P}^*\{\sup_{n > m} \|Z_n\| > \varepsilon\} < \eta$$

existiert. $v_n \xrightarrow{\text{P-f.s.}} \infty$ ist äquivalent dazu, daß für alle $\eta, m > 0$ ein $K \in \mathbb{N}$ mit

$$\mathbb{P}\{\sup_{n > K} v_n \leq m\} < \eta$$

existiert. Beides zusammen liefert für alle $\eta, \varepsilon > 0$ die Existenz von $m, M \in \mathbb{N}$, so daß

$$\begin{aligned} \mathbb{P}^*\{\sup_{n>M} \|Z_{v_n}\| > \varepsilon\} &\leq \mathbb{P}^*\{\sup_{n>M} \|Z_{v_n}\| > \varepsilon, \sup_{n>M} v_n > m\} + \mathbb{P}\{\sup_{n>M} v_n \leq m\} \\ &\leq \mathbb{P}^*\{\sup_{n>m} \|Z_n\| > \varepsilon\} + \mathbb{P}\{\sup_{n>M} v_n \leq m\} \\ &\leq 2\eta \end{aligned}$$

gilt, womit die erste Behauptung bewiesen ist.

Außerdem folgt für alle $\eta, \varepsilon > 0$ die Existenz von $n, m \in \mathbb{N}$ mit

$$\begin{aligned} \mathbb{P}^*\{\|Z_{v_n}\| > \varepsilon\} &\geq \mathbb{P}^*\{\|Z_{v_n}\| > \varepsilon, v_n > m\} + \mathbb{P}\{v_n \leq m\} \\ &\geq \mathbb{P}^*\{\sup_{1>m} \|Z_1\| > \varepsilon\} + \mathbb{P}\{v_n \leq m\} \\ &> 2\eta. \end{aligned}$$

Hiermit ist die zweite Behauptung bewiesen. □

Um zu zeigen, daß $\sqrt{v_n}(\hat{\theta}_{v_n} - \theta_0)$ dieselbe Limesverteilung wie $\sqrt{n}(\hat{\theta}_n - \theta_0)$ besitzt, reicht die Bedingung $v_n \xrightarrow{\mathbb{P}\text{-f.s.}} \infty$ nicht aus, wie folgendes Beispiel für das Kontingenztafelmodell beweist: Seien $k=1$, $N_{n1} \equiv 1$, $X_n \equiv 0$ und $\mathcal{L}(D_{n1}) = \mathcal{B}(1, \theta)$, $\forall \theta \in (0, 1)$, $n \in \mathbb{N}$. Definiere

$$v_n := \begin{cases} 1 & , \text{ falls } n=1 \\ \min\{m > v_{n-1} : \sum_{1=1}^m D_{11} = \frac{m}{2}\} & , \text{ falls } n>1 \end{cases}$$

wobei das Minimum auf der rechten Seite gleich v_{n-1} gesetzt wird, falls es nicht existiert. Dann gelten im Fall $\theta = 1/2$

$v_n \xrightarrow{\mathbb{P}\text{-f.s.}} \infty$ (Feller (68), S. 76) und

$$\mathbb{P}\{\hat{\theta}_{v_n} = \frac{1}{2}\} = \mathbb{P}\left\{\sum_{1=1}^{v_n} D_{11} = \frac{v_n}{2}\right\} = 1 \quad \forall n > 1,$$

so daß die Limesverteilung von $\hat{\theta}_{v_n}$ eine Einpunktverteilung ist.

Wir fordern deshalb zunächst zusätzlich die stochastische Unabhängigkeit der Folgen $\{v_n\}$ und $\{Y_n\}$. Um zu zeigen, daß diese Bedingungen hinreichend sind, beweisen wir als nächstes das folgende Lemma.

Lemma 13.2

Sei μ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^s, \mathcal{B}^s)$, seien $Z_1, Z_2, \dots$ s -dimensionale Zufallsvariablen mit

$$(13.1) \quad Z_n \xrightarrow{\mathcal{L}} \mu$$

und sei $\{v_n\}$ eine Folge von Zufallsvariablen in $\mathbb{N}$ mit $v_n \xrightarrow{\mathbb{P}} \infty$. Außerdem seien $\{Z_n\}$ und $\{v_n\}$ stochastisch unabhängig. Dann gilt

$$Z_{v_n} \xrightarrow{\mathcal{L}} \mu.$$

Beweis:

Für jedes $A \in \mathcal{B}^s$ gilt wegen der Unabhängigkeitsannahme und (13.1) für alle $n \in \mathbb{N}$

$$(13.2) \quad \begin{aligned} \mathbb{P}\{Z_{v_n} \in A \mid v_n = m\} &= \mathbb{P}\{Z_m \in A\} \\ &= \mu(A) + g(m) \quad \forall m \in \mathbb{N} \text{ mit } \mathbb{P}\{v_n = m\} > 0, \end{aligned}$$

wobei $g(\cdot)$ eine beschränkte, meßbare Funktion mit $\lim_{m \rightarrow \infty} g(m) = 0$ ist. Wegen $v_n \xrightarrow{\mathbb{P}} \infty$ und Lemma 6.1 gilt

$$(13.3) \quad \mathbb{E}[g(v_n)] \longrightarrow 0.$$

Aus (13.2) und (13.3) erhalten wir nun

$$\begin{aligned} \mathbb{P}\{Z_{v_n} \in A\} &= \int \mathbb{P}\{Z_{v_n} \in A \mid v_n\} d\mathbb{P} \\ &= \int [\mu(A) + g(v_n)] d\mathbb{P} \longrightarrow \mu(A), \end{aligned}$$

womit das Lemma bewiesen ist. □

Mit Hilfe dieses Lemmas läßt sich die Behauptung nun einfach zeigen :

Satz 13.3

Gelten die Voraussetzungen des Satzes 12.1, sei $\{v_n\}$ eine Folge von Zufallsvariablen in $\mathbb{N}$ mit $v_n \xrightarrow{\mathbb{P}} \infty$ und seien $\{Y_n\}$ und $\{v_n\}$ stochastisch unabhängig. Dann folgt

$$\sqrt{v_n}(\hat{\theta}_{v_n} - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, J^{-1} B J) .$$

Beweis :

Definiere

$$(13.4) \quad Z_n := \frac{1}{\sqrt{v_n}} \sum_{l=1}^{v_n} D_2 \phi(Y_l, \theta_0)', \quad n \in \mathbb{N}.$$

Dann folgt aus (12.5) und Lemma 13.2

$$Z_{v_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0, B),$$

und aus (12.6) und Lemma 13.1

$$J_{v_n} \sqrt{v_n}(\hat{\theta}_{v_n} - \theta_0) + Z_{v_n} \xrightarrow{\mathbb{P}} 0,$$

wobei $J_{v_n} \xrightarrow{\mathbb{P}} J$ gilt. Beides zusammen liefert nach gleicher Argumentation wie im Beweis zu Satz 12.1 die Behauptung. □

Im Fall, daß die Y_n unabhängig und identisch verteilt sind, gilt die Behauptung dieses Satzes auch dann, wenn von $\{v_n\}$ nur gefordert wird, daß eine Folge $\{a_n\}$ positiver Konstanten mit $a_n \rightarrow \infty$ derart existiert, daß v_n/a_n nach Wahrscheinlichkeit gegen eine positive Zufallsvariable konvergiert.

Satz 13.4

Gelten die Voraussetzungen des Satzes 12.1. Seien zusätzlich $Y_1, Y_2, \dots$ identisch verteilt und sei $\{v_n\}$ eine Folge von Zufallsvariablen in $\mathbb{N}$ mit

$$\frac{v_n}{a_n} \xrightarrow{\mathbb{P}} W,$$

wobei a_n positive Konstanten mit $a_n \rightarrow \infty$ sind und W eine positive Zufallsvariable ist. Dann folgt

$$\sqrt{v_n}(\hat{\theta}_{v_n} - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, J^{-1} B J).$$

Beweis:

Sei λ ein r -dimensionaler Spaltenvektor mit $\lambda \neq 0$. Definiere

$$\xi_n := D_2 \Phi(Y_n, \theta_0) \lambda \quad \forall n \in \mathbb{N}.$$

Dann sind die ξ_n unabhängig und identisch verteilt mit $E\{\xi_n\} = 0$ und $\text{Var}\{\xi_n\} = \lambda' B \lambda$. Hieraus folgt mit einem Theorem von Billingsley (68, S.147)

$$\frac{1}{\sqrt{v_n \lambda' B \lambda}} \sum_{i=1}^n \xi_{v_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

und weiter mit dem Satz von Cramer - Wold (Billingsley (68), S.49)

$$Z_{v_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0, B),$$

wobei Z_n durch (13.4) definiert ist. Die Behauptung folgt nun genau wie im Beweis zu Satz 13.3



Satz 13.4 ist zum Beispiel nützlich, wenn ein Experiment solange weitergeführt werden soll, bis eine vorgegebene Genauigkeit des Schätzers erreicht ist (Sequentielles Schätzen).

ANHANG A

Mannigfaltigkeit im $\mathbb{R}^k$

Wir stellen ohne Beweis einige Sätze über Mannigfaltigkeiten zusammen, die in Kapitel I benötigt werden. Die Darstellung basiert auf der von Loomis, Sternberg (68, S. 172ff) gegebenen Einführung.

Definition:

- Eine Menge $\mathcal{P} \subset \mathbb{R}^k$, $k \geq 1$, heißt r -dimensionales Pflaster (patch), wenn es einen Isomorphismus

$$\begin{aligned} \varphi(\cdot): \mathbb{R}^k &\longrightarrow \mathbb{R}^r \times \mathbb{R}^{k-r} \\ a &\longrightarrow (\varphi_1(a), \varphi_2(a)), \end{aligned}$$

$0 \leq r \leq k$, derart gibt, daß das Bild $\varphi_1(\mathcal{P})$ offen im $\mathbb{R}^r$ liegt, und wenn es eine stetige Funktion $G(\cdot): \varphi_1(\mathcal{P}) \longrightarrow \mathbb{R}^{k-r}$ gibt, so daß für alle $a \in \mathcal{P}$

$$\varphi_2(a) = G \circ \varphi_1(a)$$

gilt.

- Eine Teilmenge $\mathcal{M}$ im $\mathbb{R}^k$ heißt r -dimensionale Mannigfaltigkeit, wenn jeder Punkt aus $\mathcal{M}$ eine Umgebung besitzt, deren Schnitt mit $\mathcal{M}$ ein r -dimensionales Pflaster ist.
- Eine Mannigfaltigkeit (Pflaster) heißt C^m -Mannigfaltigkeit (C^m -Pflaster), wenn für jedes der Pflaster die zugehörige Funktion $G(\cdot)$ m -mal stetig differenzierbar ist.

Es ist sofort klar, daß jedes Pflaster eine Mannigfaltigkeit ist. Im Fall $r=0$ ist $\mathcal{M}$ genau dann eine Mannigfaltigkeit, wenn sie nur aus isolierten Punkten besteht.

Die ersten beiden Sätze geben Bedingungen an, unter denen eine durch eine stetig differenzierbare Parametrisierung $\pi(\cdot)$

gegebene Teilmenge $\mathcal{Y} \subset \mathbb{R}^k$ lokal (Satz A1) bzw. global (Satz A2) eine C^1 -Mannigfaltigkeit ist.

Satz A1

Sei $\pi(\cdot)$ eine stetig differenzierbare Abbildung von einer offenen Teilmenge $\Theta \subset \mathbb{R}^r$ nach $\mathbb{R}^k$, $0 \leq r \leq k$, und sei für alle $\theta \in \Theta$ der Rang der Ableitung $D\pi(\theta)$ gleich r . Dann besitzt jeder Punkt $\theta \in \Theta$ eine offene Umgebung U , so daß $\pi(U)$ ein C^1 -Pflaster ist.

Satz A2

Gelten dieselben Voraussetzungen wie im Satz A1. Existiert zusätzlich die Umkehrabbildung $\pi^{-1}(\cdot): \pi(\Theta) \longrightarrow \Theta$ und ist diese stetig, so ist $\pi(\Theta)$ eine r -dimensionale C^1 -Mannigfaltigkeit.

Umgekehrt ist es möglich, eine C^1 -Mannigfaltigkeit wenigstens lokal durch eine Parametrisierung zu beschreiben:

Satz A3

Sei $\mathcal{Y}$ eine r -dimensionale C^1 -Mannigfaltigkeit im $\mathbb{R}^k$. Dann gibt es zu jedem Punkt aus $\mathcal{Y}$ eine offene Umgebung U und eine stetig differenzierbare, injektive Funktion $\pi(\cdot)$ von einer offenen Teilmenge $\Theta \subset \mathbb{R}^r$ nach $\mathbb{R}^k$, deren Bild der Durchschnitt $U \cap \mathcal{Y}$ ist, deren Ableitung $D\pi(\theta)$ für alle $\theta \in \Theta$ den Rang r besitzt und deren Umkehrabbildung $\pi^{-1}(\cdot): U \cap \mathcal{Y} \longrightarrow \Theta$ stetig ist.

Der folgende Satz gibt Bedingungen dafür an, daß die als Kern einer Abbildung $R(\cdot)$ definierte Teilmenge $\mathcal{Y} \subset \mathbb{R}^k$ eine C^1 -Mannigfaltigkeit ist.

Satz A4

Sei $R(\cdot)$ eine stetig differenzierbare Abbildung von einer offenen Teilmenge $A \subset \mathbb{R}^k$ nach $\mathbb{R}^{k-r}$, $0 \leq r \leq k$. Sei $\mathcal{V}$ der Kern von $R(\cdot)$, d.h. $\mathcal{V} := \{a \in A: R(a) = 0\}$ und besitze die Ableitung $DR(a)$ für alle $a \in \mathcal{V}$ den Rang $k-r$. Dann ist $\mathcal{V}$ eine r -dimensionale C^1 -Mannigfaltigkeit.

Auch hier ist es möglich, eine C^1 -Mannigfaltigkeit wenigstens lokal als Kern einer stetig differenzierbaren Funktion $R(\cdot)$ zu beschreiben:

Satz A5

Sei $\mathcal{V}$ eine r -dimensionale C^1 -Mannigfaltigkeit im $\mathbb{R}^k$, $0 \leq r \leq k$. Dann gibt es zu jedem $a \in \mathcal{V}$ eine offene Umgebung U und eine stetig differenzierbare Funktion $R(\cdot): U \rightarrow \mathbb{R}^{k-r}$, so daß $U \cap \mathcal{V} = \{a \in U: R(a) = 0\}$ gilt und die Ableitung $DR(a)$ für alle $a \in U \cap \mathcal{V}$ den Rang $k-r$ besitzt.

Falls zwei C^1 -Mannigfaltigkeiten $\mathcal{V}$ und $\mathcal{V}'$ mit $\mathcal{V}' \subset \mathcal{V}$ vorgegeben sind, so ist folgende lokale Charakterisierung von $\mathcal{V}'$ nützlich.

Satz A6

Seien $\mathcal{V}$ und $\mathcal{V}'$ eine r - und eine s -dimensionale C^1 -Mannigfaltigkeit im $\mathbb{R}^k$, $0 \leq s \leq r \leq k$, mit $\mathcal{V}' \subset \mathcal{V}$. Seien weiter für festes $a \in \mathcal{V}'$ die offenen Mengen $U \subset \mathbb{R}^k$, $\Theta \subset \mathbb{R}^r$ und die Funktion $\pi(\cdot)$ wie in Satz A3 definiert. Dann ist $\pi^{-1}(U \cap \mathcal{V}')$ eine s -dimensionale C^1 -Mannigfaltigkeit in $\Theta \subset \mathbb{R}^r$.

Ist umgekehrt $\mathcal{V}$ durch eine Parametrisierung $\pi(\cdot): \Theta \rightarrow \mathcal{V}$ gegeben und $\mathcal{T}$ eine C^1 -Mannigfaltigkeit in Θ , dann ist auch $\pi(\mathcal{T})$ eine C^1 -Mannigfaltigkeit:

Satz A7

Sei $\pi(\cdot)$ eine stetig differenzierbare Abbildung von einer offenen Teilmenge $\Theta \subset \mathbb{R}^r$ nach $\mathbb{R}^k$, $0 \leq r \leq k$, wobei die inverse Abbildung $\pi^{-1}(\cdot): \pi(\Theta) \longrightarrow \Theta$ existiert und stetig ist und die Ableitung $D\pi(\theta)$ für alle $\theta \in \Theta$ vollen Rang r besitzt. Ist $\mathcal{T}$ eine s -dimensionale C^1 -Mannigfaltigkeit im $\mathbb{R}^r$ mit $\mathcal{T} \subset \Theta$, $0 \leq s \leq r$, so ist $\mathcal{Y}' := \pi(\mathcal{T})$ eine s -dimensionale C^1 -Mannigfaltigkeit im $\mathbb{R}^k$.

ANHANG B

Konvergenzarten bei nicht notwendig meßbaren Variablen

Um den Fall nicht meßbarer Variablen behandeln zu können, ist es erforderlich, auch nicht meßbaren Mengen Wahrscheinlichkeiten zuzuordnen. Dazu definieren wir für ein gegebenes Wahrscheinlichkeitsmaß $\mathbb{P}$ auf einem Meßraum $(\Omega, \mathcal{A})$ die äußere Wahrscheinlichkeit $\mathbb{P}^*$ bzw. innere Wahrscheinlichkeit $\mathbb{P}_*$ durch

$$\mathbb{P}^*(A) := \inf\{\mathbb{P}(C) : C \in \mathcal{A}, A \subset C\} \quad \forall A \subset \Omega$$

bzw.

$$\mathbb{P}_*(A) := \sup\{\mathbb{P}(C) : C \in \mathcal{A}, C \subset A\} \quad \forall A \subset \Omega.$$

$\mathbb{P}^*$ und $\mathbb{P}_*$ sind also für alle Teilmengen von Ω definiert. Sie hängen über

$$\mathbb{P}_*(A) = 1 - \mathbb{P}^*(\Omega - A) \quad \forall A \subset \Omega$$

zusammen. $\mathbb{P}^*$ besitzt die Eigenschaften

- (i) $\mathbb{P}^*(\emptyset) = 0$, $\mathbb{P}^*(\Omega) = 1$,
- (ii) $\mathbb{P}^*(A) \geq 0 \quad \forall A \subset \Omega$,
- (iii) $A, B \subset \Omega, A \subset B \implies \mathbb{P}^*(A) \leq \mathbb{P}^*(B)$,
- (iv) $\mathbb{P}^*\left(\bigcup_n A_n\right) \leq \sum_n \mathbb{P}^*(A_n) \quad \forall \text{ disjunkte Mengen } A_1, A_2, \dots \subset \Omega.$

Für $\mathbb{P}_*$ gelten (i) - (iii) entsprechend, während sich bei (iv) das Ungleichheitszeichen umkehrt. $\mathbb{P}^*$ und $\mathbb{P}_*$ erfüllen also bis auf die δ -Additivität alle Eigenschaften von Wahrscheinlichkeitsmaßen.

Mit Hilfe dieser äußeren Wahrscheinlichkeit verallgemeinern wir nun das Konzept der schwachen und starken Konvergenz sowie der Verteilungskonvergenz. Seien hierzu $\{\mathbb{P}_n\}$ eine Folge von Wahrscheinlichkeitsmaßen auf $(\Omega, \mathcal{A})$ und $\{X_n\}$ eine Folge von (nicht notwendig meßbaren) Variablen $\Omega \longrightarrow \mathbb{R}^k, k \in \mathbb{N}$.

Definition:

- (i) X_n konvergiert nach $\mathbb{P}_n$ -Wahrscheinlichkeit gegen eine Variable $X: \Omega \rightarrow \mathbb{R}^k$ (in Zeichen: $X_n \xrightarrow{\mathbb{P}_n} X$) : $\Longleftrightarrow \mathbb{P}_n^* \{ \|X_n - X\| > \varepsilon \} \rightarrow 0 \quad \forall \varepsilon > 0.$
- (ii) X_n konvergiert $\mathbb{P}_n$ -fast sicher gegen eine Variable $X: \Omega \rightarrow \mathbb{R}^k$ (in Zeichen: $X_n \xrightarrow{\mathbb{P}_n\text{-f.s.}} X$) : $\Longleftrightarrow \mathbb{P}_n^* \{ \sup_{m \geq n} \|X_m - X\| > \varepsilon \} \rightarrow 0 \quad \forall \varepsilon > 0.$
- (iii) X_n konvergiert nach $\mathbb{P}_n$ -Verteilung gegen ein Wahrscheinlichkeitsmaß μ auf $(\mathbb{R}^k, \mathcal{B}^k)$ (in Zeichen: $X_n \xrightarrow{\mathcal{L}_n} \mu$) : $\Longleftrightarrow \mathbb{P}_n^* \{ X_n \in B \} \rightarrow \mu(B) \quad \forall B \in \mathcal{B}^k \text{ mit } \mu(\partial B) = 0. \quad +)$

Für $X_n \xrightarrow{\mathcal{L}_n} \mu$ schreiben wir auch äquivalent $\mathcal{L}_n(X_n) \rightarrow \mathcal{L}(X)$, wobei X eine Zufallsvariable nach $(\mathbb{R}^k, \mathcal{B}^k)$ ist, die die Verteilung μ besitzt.

In dem nachfolgenden Satz sind alle die in dieser Arbeit benötigten Eigenschaften bzgl. dieser Konvergenzarten zusammengestellt.

Satz

Seien X_n und Y_n (nicht notwendig meßbare) Variablen $\Omega \rightarrow \mathbb{R}^k$, $k \in \mathbb{N}$. Weiter seien X eine Zufallsvariable $(\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^k, \mathcal{B}^k)$, $a = (a_1, \dots, a_k)'$ ein beliebiger Punkt aus $\mathbb{R}^k$ und $f(\cdot)$ eine Funktion $\mathbb{R}^k \rightarrow \mathbb{R}^1$, $l \in \mathbb{N}$. Dann gelten

- (i) $X_n \xrightarrow{\mathbb{P}_n\text{-f.s.}} X \implies X_n \xrightarrow{\mathbb{P}_n} X,$
- (ii) $X_n \xrightarrow{\mathbb{P}_n} a \iff \mathcal{L}_n(X_n) \rightarrow \mathcal{L}(a),$
- (iii) $X_n \xrightarrow{\mathbb{P}_n} a \iff X_{ni} \xrightarrow{\mathbb{P}_n} a_i, \quad i=1, \dots, k,$

+) ∂B bezeichnet den Rand von B .

$$\begin{aligned} \text{(iv)} \quad \mathcal{L}_n(X_n) \longrightarrow \mathcal{L}(X), Y_n \xrightarrow{\mathbb{P}_n} a \implies \mathcal{L}_n(X_n + Y_n) \longrightarrow \mathcal{L}(X + a), \\ Y_n' X_n - a' X_n \xrightarrow{\mathbb{P}_n} 0, \\ \mathcal{L}_n(Y_n' X_n) \longrightarrow \mathcal{L}(a' X). \end{aligned}$$

$$\text{(v)} \quad \mathcal{L}_n(X_n) \longrightarrow \mathcal{L}(X), \mu^*(S) = 0 \text{ (S ist die Menge der Unstetigkeitsstellen von } f(\cdot) \text{)} \implies \mathcal{L}_n\{f(X_n)\} \longrightarrow \mathcal{L}\{f(X)\}.$$

$$\begin{aligned} \text{(vi)} \quad X_n \xrightarrow{\mathbb{P}_n} a, f(\cdot) \text{ stetig differenzierbar in } a \\ \implies f(X_n) = f(a) + Df(a)(X_n - a) + \|X_n - a\| o_{\mathbb{P}_n}(1). \end{aligned}$$

Beweis :

(i) folgt sofort aus den Definitionen.

(ii) $\mathcal{L}_n(X_n) \longrightarrow \mathcal{L}(a)$ ist äquivalent zu der Beziehung

$$\mathbb{P}_n^* \{X_n \in B\} \longrightarrow I_B(a) \quad \forall B \in \mathcal{B}^k \text{ mit } a \notin \partial B,$$

welche wiederum äquivalent ist zu

$$\mathbb{P}_n^* \{\|X_n - a\| > \varepsilon\} \longrightarrow 0 \quad \forall \varepsilon > 0.$$

(iii) erhält man leicht aus

$$\begin{aligned} \mathbb{P}_n^* \{|X_{ni} - a_i| > \varepsilon\} &\leq \mathbb{P}_n^* \{\|X_n - a\| > \varepsilon\} \\ &\leq \sum_{j=1}^k \mathbb{P}_n^* \{|X_{nj} - a_j| > \varepsilon\} \quad \forall \varepsilon > 0, i=1, \dots, k. \end{aligned}$$

(iv) Sei $B \in \mathcal{B}^k$ mit $\mathbb{P}\{X+a \in \partial B\} = 0$. Seien weiter $B_\varepsilon := \{x \in \mathbb{R}^k : \|x-y\| < \varepsilon \text{ für ein } y \in B\}$ und $B^\varepsilon := \mathbb{R}^k - (B^k - B)_\varepsilon$. Da es höchstens abzählbar viele verschiedene $\varepsilon > 0$ mit $\mathbb{P}\{X+a \in \partial B_\varepsilon\} > 0$ oder $\mathbb{P}\{X+a \in \partial B^\varepsilon\} > 0$ gibt, existiert eine Folge $\{\varepsilon_m\}$ positiver Zahlen mit $\varepsilon_m \rightarrow 0$ und $\mathbb{P}\{X+a \in \partial B_{\varepsilon_m}\} = 0$ und $\mathbb{P}\{X+a \in \partial B_{\varepsilon_m}^c\} = 0 \quad \forall m \in \mathbb{N}$. Somit gilt für jedes $m \in \mathbb{N}$ nach Voraussetzung $\mathbb{P}_n^* \{X_n + a \in B_{\varepsilon_m}^c\} \rightarrow \mathbb{P}\{X+a \in B_{\varepsilon_m}^c\}$ und $\mathbb{P}_n^* \{X_n + a \in B_{\varepsilon_m}\} \rightarrow \mathbb{P}\{X+a \in B_{\varepsilon_m}\}$, und wir erhalten mit $Y_n \xrightarrow{\mathbb{P}_n} a$:

$$\begin{aligned} \mathbb{P}_n^* \{X_n + Y_n \in B\} &\leq \mathbb{P}_n^* \{X_n + Y_n \in B, \|Y_n - a\| \leq \varepsilon_m\} + \mathbb{P}_n^* \{\|Y_n - a\| > \varepsilon_m\} \\ &\leq \mathbb{P}_n^* \{X_n + a \in B_{\varepsilon_m}\} + \mathbb{P}_n^* \{\|Y_n - a\| > \varepsilon_m\} \\ &\longrightarrow \mathbb{P}\{X + a \in B_{\varepsilon_m}\} \quad \forall m \in \mathbb{N} \end{aligned}$$

und

$$\begin{aligned} \mathbb{P}_n^* \{X_n + Y_n \in B\} &\geq \mathbb{P}_n^* \{X_n + Y_n \in B, \|Y_n - a\| \leq \varepsilon_m\} \\ &\geq \mathbb{P}_n^* \{X_n + a \in B_{\varepsilon_m}\} \\ &\longrightarrow \mathbb{P}\{X + a \in B_{\varepsilon_m}\} \quad \forall m \in \mathbb{N}. \end{aligned}$$

Da wegen $\mathbb{P}\{X + a \in \partial B\} = 0$ $\lim_{m \rightarrow \infty} \mathbb{P}\{X + a \in B_{\varepsilon_m}\} = \mathbb{P}\{X + a \in B\}$ und $\lim_{m \rightarrow \infty} \mathbb{P}\{X + a \in B_{\varepsilon_m}\} = \mathbb{P}\{X + a \in B\}$ gelten, ist damit die erste Behauptung bewiesen. Weiterhin gilt für alle $\varepsilon > 0$ und alle $K > 0$ mit $\mathbb{P}\{\|X\| = K\} = 0$

$$\begin{aligned} \mathbb{P}_n^* \{|(Y_n - a)' X_n| > \varepsilon\} &\leq \mathbb{P}_n^* \{|(Y_n - a)' X_n| > \varepsilon, \|Y_n - a\| \leq \frac{\varepsilon}{K}\} + \mathbb{P}_n^* \{\|Y_n - a\| > \frac{\varepsilon}{K}\} \\ &\leq \mathbb{P}_n^* \{\|X_n\| > K\} + \mathbb{P}_n^* \{\|Y_n - a\| > \frac{\varepsilon}{K}\} \\ &\longrightarrow \mathbb{P}\{\|X\| > K\}. \end{aligned}$$

Die zweite Behauptung folgt nun, da $\lim_{K \rightarrow \infty} \mathbb{P}\{\|X\| > K\} = 0$ gilt und da man beliebig große K mit $\mathbb{P}\{\|X\| = K\} = 0$ finden kann. Die dritte Behauptung erhält man sofort aus den ersten beiden Behauptungen.

(v) Zu zeigen ist

$$\mathbb{P}_n^* \{f(X_n) \in A\} \longrightarrow \mathbb{P}\{f(X) \in A\} \quad \forall A \in \mathcal{B}^1 \text{ mit } \mathbb{P}\{f(X) \in \partial A\} = 0$$

oder äquivalent

$$\mathbb{P}_n^* \{X_n \in B\} \longrightarrow \mu(B) \quad \forall B \in f^{-1}(\mathcal{B}^1) \text{ mit } \mu(f^{-1} \partial f(B)) = 0.$$

Diese Beziehung ist erfüllt, wenn $\mu(\partial B) = 0$ aus $\mu(f^{-1} \partial f(B)) = 0$ folgt. Diese Implikation gilt jedoch wegen

$$(\mathbb{R}^k - S) \cap \partial B \subset f^{-1} \partial f(B).$$

(vi) Aus der stetigen Differenzierbarkeit von $f(\cdot)$ im Punkt a folgt

$$f(x) = f(a) + Df(a)(x - a) + \|x - a\| g(x), \quad x \in \mathbb{R}^k,$$

wobei $g(\cdot)$ eine Funktion mit $\lim_{x \rightarrow a} g(x) = 0$ ist. Somit ist $g(\cdot)$ im Punkt a stetig und wir erhalten mit (v)

$$g(X_n) \xrightarrow{P_n} g(a) = 0$$

und damit die Behauptung.



Bezeichnungen

$\ \cdot\ $	Maximum - Norm (bei Vektoren und Matrizen)
$I_M(x)$	$:= \begin{cases} 1, & \text{falls } x \in M \\ 0, & \text{sonst} \end{cases}$, Indikatorfunktion
$\text{diag}\{a_i\}$	Diagonalmatrix mit Komponente a_i an i -ter Stelle in der Diagonalen
I	$:= \text{diag}\{1\}$ Einheitsmatrix
$Df(u)$	Ableitung der Funktion $f(\cdot): \mathbb{R}^J \rightarrow \mathbb{R}^I$ an der Stelle u . Für $J > 0$: $I \times J$ -Matrix der partiellen Ableitungen. Für $J=0$: I -Spaltenvektor, bestehend aus lauter Nullen.
$D_m f(u_1, \dots, u_K)$	partielle Ableitung der Funktion $f(\cdot): \mathbb{R}^{J_1} \times \dots \times \mathbb{R}^{J_K} \rightarrow \mathbb{R}^I$, $J_1, \dots, J_K, I > 0$, nach dem m -ten Argument, $1 \leq m \leq K$ ($I \times J_m$ - Matrix)
$D_m^2 f(u_1, \dots, u_K)$	$J_m \times J_m$ - Matrix der zweiten Ableitung von $f(\cdot): \mathbb{R}^{J_1} \times \dots \times \mathbb{R}^{J_K} \rightarrow \mathbb{R}$, $J_1, \dots, J_K > 0$, nach dem m -ten Argument, $1 \leq m \leq K$
$\mathcal{M}(A)$	linearer Teilraum, der von den Spalten der Matrix A aufgespannt wird.
P_A	orthogonale Projektion auf $\mathcal{M}(A)$ als Matrix
$f \circ g(\cdot)$	$:= f(g(\cdot))$ Komposition von Abbildungen
$f^{(c)}(\cdot)$	$:= \min_{c \in C} \{f(\cdot)\}$, $c \in \mathbb{R}$
$\mathcal{B}(m, p)$	Binomialverteilung mit den Parametern $m \in \mathbb{N}_0$, $p \in [0, 1]$. $\mathcal{B}(0, p)$ sei als Verteilung mit gesamter Masse auf $\{0\}$ definiert.
$\mathcal{M}(m, p)$	Multinomialverteilung mit den Parametern $m \in \mathbb{N}_0$, $p \in [0, 1]^t$, $t > 0$
$\mathcal{N}(a, \Sigma)$	multidimensionale Normalverteilung mit Erwartungsvektor a und Kovarianzmatrix Σ

$\chi^2_t(\delta)$	nichtzentrale χ^2 -Verteilung mit t Freiheitsgraden und Nichtzentralitätsparameter δ
$\Phi(\cdot)$	Verteilungsfunktion der eindimensionalen standardisierten Normalverteilung
$E_a\{X\}, \text{Var}_a\{X\}, \mathcal{L}_a(X)$	Erwartungswert, Varianz, Verteilung der Zufallsvariablen X unter dem Wahrscheinlichkeitsmaß P_a
$P_a(A Y), E_a\{X Y\}$ $[P_a(A \mathcal{T}), E_a\{X \mathcal{T}\}]$	bedingte Wahrscheinlichkeit von A , bedingter Erwartungswert von X bedingt auf Y ($\mathcal{G}$ -Algebra $\mathcal{T}$)
$P_a(\cdot Y) [P_a(\cdot \mathcal{T})]$	bedingte Verteilung von P_a gegeben Y ($\mathcal{G}$ -Algebra $\mathcal{T}$)
$\mathcal{L}_a(X Y=y)$	$:= P_a\{X \in \cdot Y=y\}$ bedingte Verteilung von X unter P_a gegeben $Y=y$
P^*, P_*	äußeres, inneres Maß zu P (siehe Anhang B)
$\mathcal{G}^S$	s -dimensionale Borelsche $\mathcal{G}$ -Algebra auf $\mathbb{R}^S$ (oder einer Teilmenge des $\mathbb{R}^S$)
$\xrightarrow{P_n}$	Konvergenz nach P_n -Wahrscheinlichkeit (siehe Anhang B)
$\xrightarrow{P\text{-f.s.}}$	P -fast sichere Konvergenz (siehe Anhang B)
$o_{P_n}(a_n)$	$X_n = o_{P_n}(a_n) : \iff X_n/a_n \xrightarrow{P_n} 0 \ (n \rightarrow \infty)$
$O_{P_n}(a_n)$	$X_n = O_{P_n}(a_n) : \iff$ für alle $\eta > 0$ gibt es eine Konstante $K \in \mathbb{R}$, so daß $P_n^*\{\ X_n\ /a_n > K\} < \eta \quad \forall n \in \mathbb{N}$ gilt
$\xrightarrow{\mathcal{L}_n}$	Konvergenz nach Verteilung unter P_n (siehe Anhang B)
N_0	$:= \mathbb{N} \cup \{0\}$

Literaturverzeichnis

- Ashton, W.D. (1972). The logit transformation with special reference to its uses in bioassay. Charles Griffin & Co., London.
- Billingsley, P. (1968). Convergence of probability measures. John Wiley & Sons, New York.
- Cox, D.R. (1970). Analysis of binary data. Methuan, London.
- Cox, D.R. (1975). Partial likelihood. Biometrika 62, S.269 - 276.
- Feller, W. (1968). An introduction to probability theory and its applications. Vol. 1, 3rd edition. John Wiley & Sons, New York.
- Finney, D.J. (1971). Probit analysis. 3rd edition. Cambridge University Press, Cambridge.
- Gänssler, P. und Stute, W. (1977). Wahrscheinlichkeitstheorie. Springer - Verlag, Berlin.
- Haberman, S.J. (1974). The analysis of frequency data. The University of Chicago Press, Chicago.
- Hoadley, B. (1971). Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case. Ann. Math. Statist. 42, S.1977 - 1991.
- Kalbfleisch, J.D. and Prentice, R.L. (1980). The statistical analysis of failure time data. John Wiley & Sons, New York.
- Linder, A. und Berchtold, W. (1976). Statistische Auswertung von Prozentzahlen. UTB 522. Birkhäuser Verlag, Basel.
- Loève, M. (1977). Probability theory. Vol. 1, 4th edition. Springer - Verlag, New York.
- Loève, M. (1977a). Probability theory. Vol. 2, 4th edition. Springer - Verlag, New York.
- Loomis, L.H. and Sternberg, S. (1968). Advanced calculus. Addison - Wesley Publishing Company, Reading (Mass.).
- Rao, C.R. (1963). Criteria of estimation in large samples. Sankhyā A 25, S.189 - 206.

- Rao, C.R. (1973). Linear statistical inference and its applications. 2nd edition. John Wiley & Sons, New York.
- Schumacher, M. (1978). Schätzmethoden in der Dosis - Wirkungs - Analyse. Arbeitspapier 78/13 der Abteilung Statistik, Universität Dortmund.
- Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. Ann.Math. Statist. 20, S.595 - 601.
- Witting, H. und Nölle, G. (1970). Angewandte Mathematische Statistik. BiG. Teubner, Stuttgart.

